



CITATION

Sileo, A. AI-Driven Dubbing in Italy: Taking Stock of Local Audience Reception. *Linguae & - Journal of Modern Languages and Cultures*, 29(1). <https://doi.org/10.14276/l.v29i1.5550>

DOI

<https://doi.org/10.14276/l.v29i1.5550>

RECEIVED

March 1, 2026

ACCEPTED

October 9, 2026

PUBLISHED

October 17, 2026

PEER REVIEW HISTORY

double blind review

COPYRIGHT



This is an open access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.



AI-Driven Dubbing in Italy: Taking Stock of Local Audience Reception

Angela Sileo (San Raffaele Open University (Rome))

angela.sileo@uniroma5.it

ABSTRACT

This study investigates the responses of young Italian audiences to AI-generated dubbing in Audiovisual Translation (AVT). By surveying three distinct age groups (15-19, 20-24, and 25-34), the research examines the acoustic naturalness, linguistic quality, willingness to pay, and genre preferences. Results indicate a significant generational divide heavily influenced by digital media consumption. While younger digital natives (15-19) demonstrate a higher susceptibility to the auditory illusion of AI – likely due to their exposure to synthetic voices on social media –, the 20-24 demographic emerges as highly critical, firmly rejecting the AI's unnatural linguistic choices and phrasing. Economically, the study reveals a unanimous rejection across all groups toward paying cinema tickets for AI-dubbed films. However, audiences show a strong willingness to consume free, AI-dubbed content at home, predominantly restricting their acceptance to unscripted formats. Ultimately, while AI is slowly being accepted as a cost-effective, supplementary tool for secondary audiovisual products, audiences still fiercely protect the emotional authenticity of cinematic storytelling, rejecting the hypothesis that artificial intelligence will eventually and entirely replace traditional, human-led dubbing.

Angela Sileo

San Raffaele Open University (Rome)
angela.sileo@uniroma5.it

AI-Driven Dubbing in Italy: Taking Stock of Local Audience Reception

ABSTRACT

This study investigates the responses of young Italian audiences to AI-generated dubbing in Audiovisual Translation (AVT). By surveying three distinct age groups (15-19, 20-24, and 25-34), the research examines the acoustic naturalness, linguistic quality, willingness to pay, and genre preferences. Results indicate a significant generational divide heavily influenced by digital media consumption. While younger digital natives (15-19) demonstrate a higher susceptibility to the auditory illusion of AI – likely due to their exposure to synthetic voices on social media –, the 20-24 demographic emerges as highly critical, firmly rejecting the AI's unnatural linguistic choices and phrasing. Economically, the study reveals a unanimous rejection across all groups toward paying cinema tickets for AI-dubbed films. However, audiences show a strong willingness to consume free, AI-dubbed content at home, predominantly restricting their acceptance to unscripted formats. Ultimately, while AI is slowly being accepted as a cost-effective, supplementary tool for secondary audiovisual products, audiences still fiercely protect the emotional authenticity of cinematic storytelling, rejecting the hypothesis that artificial intelligence will eventually and entirely replace traditional, human-led dubbing.

KEYWORDS: AVT; AI-driven dubbing; studio-based dubbing; Italian scenario; audience reception

1. Introduction

The introduction of artificial intelligence into the realm of AVT, particularly within the Italian context, is perceived by industry insiders – ranging from script adapters to voice actors – as a significant risk factor. Some even view it as a ‘pandemic’ that is likely to alter traditional workflows, causing ‘casualties’ within the sector.

Consequently, the local perspective on adopting AI to make audiovisual products accessible seems to lack the positive, proactive view held by those who see technology as an ally for inclusion; rather, it is viewed with profound apprehension.

On a global scale, some nations are taking steps to protect human voices and faces from exploitation, with Denmark serving as a notable example (Bryant 2025). While deepfake and swapping technologies are widely diffused and trendy on social media, they remain less common in television, likely due to the latter's more traditionalist nature in Italy. A well-known local exception is the furniture brand "ArredissimA" (Tuveri 2025) which has resorted to the sampled voice and face of the famous Italian showman Gerry Scotti to advertise its products on TV. Although the image superposition is not always flawless and the sampled voices can sound slightly robotic, improvements are constant, and every passing moment brings the public closer to a possible complete acceptance of AI-processed content. This aligns with a broader universal debate regarding the survival of human creativity, extending to acting itself – see, as an instance, the controversy surrounding AI-generated actress Tilli Norwood which has been enraging Hollywood (Goodfellow 2025) – and to journalism, with the creation of fully AI-generated news anchors (NDTV 2025).

From the Italian perspective, a strong element of novelty was introduced in the latest National Collective Labor Agreement (Contratto Collettivo Nazionale del Lavoro or CCNL) for the dubbing sector, finalized in late 2023 (AIDAC 2023). For the first time, this agreement included a specific clause, Article 22, dedicated to artificial intelligence. This article does not intend to halt technological progress but recognizes the need to safeguard the creative aspects of the work and related rights, particularly concerning the transfer of authorship and performer rights in the absence of explicit consent. Despite this, shortly after the CCNL was issued, a local broadcast on a major RAI channel used AI to clone the voice of the late Claudio Capone, a historic voice actor known for dubbing Luke Skywalker, among others. His son officially warned RAI against using his father's voice, demanding public apologies and the immediate deletion of any derivative recordings to prevent their future use for commercial or entertainment purposes (Monaco 2025). This underscores to what extent, in the Italian audiovisual landscape, the protection of vocal identity is perceived as a highly sensitive issue, a crucial one for audience reception.

Against this backdrop, the present study investigates the reception of AI-generated dubbing among young Italian audiences (aged 15-34), as younger generations are supposedly more favorable to emerging technological shifts in AVT. Its empirical focus is the 2024 edition of *Ballando con le stelle* (2005-) on RAI 1, the Italian adaptation of *Strictly Come Dancing* (2004-), which constitutes the first documented case of AI-driven dubbing broadcast on Italian prime-time national television: the production employed AI to sample the face and voice of the Turkish actor Furkan Palali in order to render his pre-recorded interviews in Italian. Drawing on a web-based questionnaire administered in late 2025 to 142 respondents from Northern, Central, and Southern Italy – segmented into three age brackets (15-19, 20-24, 25-34) –, the study addresses four interconnected research questions. Firstly, it explores whether AI-generated dubbing is likely to be accepted by the local audience (RQ1) and to eventually replace traditional studio-recorded dubbing (RQ2). Furthermore, the study examines whether acceptance depends on the genre of the audiovisual product (RQ3) and, finally, if economic factors play a role in this process (RQ4).

In line with the working hypotheses detailed in Section 4, the study anticipates a gradual and conditional acceptance of AI-driven dubbing, more pronounced among younger viewers and largely restricted to freely available, unscripted content. Three dimensions are analyzed: the perceived naturalness of the synthetic voice, the quality of the AI-generated translation, and the acceptability of the overall dubbing output.

The article is structured as follows. Section 2 situates the study within research on audience reception and introduces the notion of “tolerance threshold” (Spiteri Miggiani 2025). Section 3 reconstructs the trajectory of AVT practices in the Italian context and presents a linguistic analysis of the AI-dubbed segments under examination. Section 4 describes the survey design, the research questions and hypotheses, and the sample. Section 5 reports the results and Section 6 discusses them considering the research questions. Finally, Section 7 draws conclusions and outlines directions for future research.

2. Audience Reception and “Tolerance Threshold”

Over the past few decades, reception research has moved to the center of academic inquiry, a shift prompted by the recognition that translation choices can only be fully evaluated by considering how actual viewers respond to them (Di Giovanni and Gambier 2018). Within this dynamic field – which evolves alongside technological advancements and increasingly embraces accessibility awareness –, a foundational distinction is drawn between viewers’ *response* (perceptual decoding, such as subtitle legibility), *reaction* (psycho-cognitive dimensions, including readability and processing effort), and *repercussion* (pertaining to attitudinal preferences, habits, and broader socio-cultural values) (Kovačič 1995; Gambier 2018). The present study is situated primarily at the level of *reaction*, while also capturing elements of *repercussion* by accounting for the potential impact of AI-dubbed media on the audience’s long-term viewing habits and audiovisual consumption styles, which are investigated using a survey instrument detailed in Section 4.

Within the complex audiovisual sphere, the voice and its presumed naturalness remain the deciding factors in the success of an AI-localized product. As highlighted by Spiteri Miggiani (2025), to ensure a friendly and comfortable viewing experience and not exceed the “tolerance threshold” of both the audience and the client, it is necessary to preserve the creativity inherent in the dubbing process. However, the audience’s acceptance threshold depends largely on familiarity and habituation (*ibid.*). Just as the local audience gradually accustomed itself to the artificial register of traditional dubbing language – which lacks real correspondence within the broader framework of standard Italian varieties –, they may eventually develop familiarity with AI-driven dubbing characterized by minimal creativity.

The critical danger lies in triggering a reaction from viewers by making the presence of a “triple duplicity” obvious. Dubbing is naturally characterized by a “double duplicity” (Sileo 2024, 293): a visual duplicity (presenting a scenic reality as factual) and a linguistic duplicity (characters moving in a specific cultural space, like a Japanese city, but speaking Italian). The Italian public is implicitly ready to accept this double layer due to habituation. However, burdening the visual experience with a third layer of duplicity – understood as a lack of natural-

ness and authenticity both in visual and in acoustic terms – risks breaching the “tolerance threshold.” This could betray the artificial nature of the service. Consequently, the viewing process may be disturbed and disrupted by imperfectly natural voices or AI’s inability to reproduce native intonation. Furthermore, linguistic issues remain, such as AI interpreting pauses as sentence boundaries, leading to grammatical disagreements between nouns and adjectives (Di Francesco and Sileo 2025).

A pivotal case study marking a turning point in the Italian sector is the 2024 edition of the TV show *Ballando con le stelle* (2005-), a localized version of the UK show *Strictly Come Dancing* (2004-). Over the show’s 20-year history, various strategies have been employed to adapt the speech of foreign contestants, typically evolving from voice-overs to simul sync-like techniques to ensure greater verisimilitude (ibid.). The 2024 edition represented a watershed, as it was the first time prime-time national television utilized AI to sample the face and voice of a contestant, Turkish actor Furkan Palali. The technology was used to translate pre-recorded interviews into Italian, whereas during live segments, the actor struggled to express himself in Italian while relying on an earpiece for simultaneous interpretation.

3. The Case in Point

To fully grasp the relevance of this phenomenon within the Italian context, it is necessary to provide a chronological overview, however brief, regarding the emergence and fortune of the main audiovisual modes implemented in Italy since the advent of sound cinema. Initially, subtitles were the primary means to localize audiovisual products within national borders until the arrival of dubbing in the early 1930s, which drastically changed the landscape. Even more so, with the Royal Decree of October 5, 1933, through which the Mussolini Government mandated that the entire dubbing process be performed within the national territory (Raffaelli 1983, 1991).

Since then, historical motivations have intertwined with cultural ones, given the high illiteracy rates of the Italian population at the time, which would have undoubtedly alienated the public from subtitled cinema. Furthermore, there

were significant linguistic and political motivations: only a few decades after the unification of Italy (1861), the nation struggled to affirm a national language. Dubbing, and later television, acted as teachers of standard Italian, contributing significantly to its diffusion (Raffaelli 1991). For all these reasons, dubbing has retained a privileged position among the average Italian audience (Sileo 2018a, 2020). This preference persists today, although new demographic segments are promoting other forms of audiovisual translation; for instance, younger audiences often prefer original audio with either L1 or L2 subtitles.

One cannot overlook thematic channels, OTT platforms, and minor television channels that, unable to bear the high costs of the dubbing supply chain, have contributed over time to the development of a translation mode often considered in industry discourse low-quality or niche. This mode is particularly suited to hybrid television genres where the flat delivery of traditional voice-over, however economical, could not offer the emotional engagement required. This evolved mode, which emerged in the mid-2000s, is known among industry insiders as “simil sync” but defined in the latest CCNL as “linear sync.”

This modality arises from a spontaneous process of “dubbingization” of voice-over (Di Francesco and Sileo 2025). Simil sync reproduces the same pauses, gaps, and sounds of the original, offering complete coverage of the original line, which remains slightly audible for the target audience to hear. It functions well with prevalent non-scripted television genres – e.g., reality shows or interviews – that are spontaneous or presumably so. While coexisting peacefully with other modes, dubbing continues to be the preferred choice, especially for the adult population and feature films, as reported by Sileo (2018a, 2020). In other words, the reasons for the predilection for dubbing over more cost-effective and less time-consuming modes are deeply cultural, historical, and linguistic. The local audience is strongly tethered to this form of translation, and the “habituation” discussed by Spiteri Miggiani (2025) is the very factor determining a sort of immobility in general taste.

However, towards the end of 2024, a situation occurred that may unsettle this consolidated equilibrium. While the 2010 edition of *Ballando con le stelle* opted for an adaptation very close to simil sync to voice American actor Ron Moss, and the following year used a mode even closer to simil sync for German actor Gedeon Burkhard, the 2024 edition marked a significant shift

when AI-generated dubbing was implemented. According to internal production sources, this stemmed from a strategic decision by the program's artistic director, who favored the technology under the premise that automated synthesis provided a more natural and credible acoustic delivery, making it highly preferable over conventional methods.

A linguistic and translational analysis of the target text seems to reveal interferences or “transfers” (Toury 1995) from English, suggesting both that the actor likely used it as a lingua franca and that a word-for-word machine translation approach was employed, coupled with an overly formal register that heightens artificiality – “mi accarezzava dolcemente la schiena” in (1) – and to a surprising lack of paralinguistic features – fillers, hesitations, false starts – that are in fact abundant in natural speech. It also features grammatical errors, such as the failure of noun-adjective agreement (using the masculine “premuroso” referring to the actor's mother in the first example provided). The AI translator also struggled with English modal verbs (*can, may, might*) in example number 4. These and other interferences detected in the contestant's speech – and illustrated in the following examples – fall into the category of *doppiaggese*, a pseudo-colloquial variety of Italian produced by mechanical translation and sustained by prolonged audience exposure (Sileo 2018b; see also Pavesi 2008; Bucaria and Chiaro 2007).

1. Sono cresciuto con il supporto di mia madre. È amorevole, e affettuosa, e premuroso, che *mi accarezzava dolcemente sempre la schiena* [*Ballando con le stelle*, 12/10/2024]

[Back Translation¹: I grew up with the support of my mother. She is lovely, and affectionate, and caring, who *always used to sweetly caress my back*]

2. [...] che entrambi mi abbiano dato *la chiave per essere una brava persona* [*Ballando con le stelle*, 12/10/2024]

[BT: that both gave me *the key to being* a good person]

¹ It is necessary to clarify that the backtranslation provided here is merely a hypothesis of what the original text might have been. It can be reasonably assumed that the original speech was in English since, during live shows, the Turkish contestant frequently exhibited instances of involuntary code-mixing between English and Italian. This leads to the assumption that English was used whenever possible, as it is a more accessible lingua franca in an Italian television setting compared to Turkish. Given this premise, it is possible to trace back to a hypothetical Ur-speech, albeit with some flexibility.

3. [...] la mia donna ideale ... che non cerca di cambiarmi, che *può* ascoltarmi senza giudicare i miei errori [Ballando con le stelle, 28/09/2024]
[BT: my ideal woman ... who doesn't try to change me, who *can* listen to me without judging my mistakes]
4. *Possono* criticare la mia danza, *potrebbero* non gradire, *potrebbero* pensare che non sia elastico [Ballando con le stelle, 12/10/2024]
[BT: They *may* criticize my dancing, they *might as well* not appreciate, they *might (as well)* think I'm not nimble]

When traces of the source text and language remain visible in the target text, it indicates that the message was not sufficiently reshaped and did not undergo an intermediate stage of human-driven re-coding of its deep structure. In other words, while the adaptation offered by artificial intelligence may be impeccable from the perspective of qualitative sync – specifically regarding lip movements, which strongly sustain the illusion of truth –, it simultaneously offers speech that is at times grammatically incorrect as well as unnatural. The mentioned issue with *premuroso* in reference to the speaker's mother is a prime example. This occurs because, as stated by Di Francesco and Sileo (2025), artificial intelligence seems to struggle with human pauses, often interpreting them as absolute sentence boundaries and thus generating a new utterance disconnected from the previous one.

Furthermore, the *doppiaggese* observed in these broadcasts seems to be heavily characterized by English calques. These can be structural when they perfectly mirror the source text's syntax, as seen in (2) with the phrase *la chiave per essere* ("the key to being"), which is an inter-syntagmatic structural calque following Sileo's (2018b) framework. However, *doppiaggese* is also characterized by semantic calques, which are less frequent but more insidious. We can observe these semantic calques in (3) and (4), illustrating the well-known translational dilemma surrounding the English verb "can" and its related modals "may" and "might." This represents a *sui generis* semantic calque, which typically occurs when an existing word in the target language acquires a new meaning due to a superficial equivalence (such as the Italian "realizzare" adopting the English meaning of "to realize/become aware"). In the case of these modal verbs, the calque is based on semantic rather than morphological similarity. Because these modals lack a strict one-to-one relationship with their Italian equivalents, a machine translation fails

at semantic disambiguation, unable to recognize the contextual difference between “knowing how to do something” (*sapere*) and “being able or permitted to do something” (*potere*). In (4), the inherent sense of concession embedded in the hypothetical source text shifts entirely in the Italian translation, improperly taking on connotations of mere probability.

Poor translation quality, stemming from a lack of creativity, is not the only factor in evaluating the acceptability of (AI) dubbing. Audience reception relies heavily on synchronization and audiovisual consistency; for example, a character shaking their head in denial should not verbally affirm a statement. The naturalness of the voice is also paramount. It is precisely the co-presence of these dimensions – voice naturalness, translation quality, and the acceptability of the overall dubbing output – that, in the absence of the “double creativity” inherent in studio dubbing (Spiteri Miggiani 2025), risks producing the “triple duplicity” discussed in Section 2. The audience holds sway and will determine the success or failure of this phenomenon; while academic roundtables and debates with professionals are valuable, whether and to what extent viewers perceive and accept these layers cannot be established through such discussions alone: it requires empirical audience research. The survey presented in the next section was designed to address this question.

4. The Survey

Attempting to comprehensively map the local reception and appreciation of specific translation modes is an ambitious endeavor that inevitably encounters certain limitations. Chief among these is the practical difficulty of identifying a sample genuinely representative of Italian public taste; nevertheless, the attempt must be made. For this reason, since 2019, the present author has aimed to disseminate quantitative surveys to gauge the reception of the Italian public. The initial survey (Sileo 2020) targeted primarily professionals, industry experts, enthusiasts, and translators. It was followed by a subsequent survey (Sileo 2025) to verify any changes in local attitudes toward the practice known as *simil sync*, as discussed in the previous section. The research questions for these earlier studies focused on the comparison between *simil sync* and traditional dubbing, seeking

to formulate predictions regarding the future of the Italian audiovisual landscape. This included questioning whether *simil sync* might eventually replace dubbing, the relevance of economic factors, and whether distinctions should be made between textual or audiovisual genres. The present study builds on this line of research, adapting the design of these earlier instruments to the case of AI-generated dubbing.

As anticipated, the 2024 edition of *Ballando con le stelle* introduced a significant novelty into the Italian audiovisual panorama. Although it remains uncertain whether this experimentation with AI marks the beginning of a long-term trend or will remain an isolated incident, it is reasonable to assert that the audience – however often a passive recipient of top-down decisions – possesses the capacity to influence the future success or failure of AI-generated dubbing; in AVT matters, what is economically advantageous for the industry and acceptable to the public generally succeeds. The study presented here relies on a web-based survey administered in late 2025.

The core research questions investigate whether AI-generated dubbing will ever be genuinely embraced by the local audience (RQ1) and eventually replace traditional studio-recorded dubbing (RQ2). The study also examines whether acceptance depends on the specific genre of the audiovisual product (RQ3), and if economic motivations may play a crucial role in this phenomenon (RQ4). The working hypotheses suggest that artificial dubbing will be slowly accepted within the local audiovisual landscape (RH1), possibly replacing traditional studio dubbing initially among younger audiences (RH2), extending even to scripted content such as films (RH3), provided that this content is offered free of charge (RH4). The survey is methodologically divided into three distinct sets of questions. The first block addresses the socio-cultural background and viewing habits of the respondents, collecting data on age, education level, area of expertise, and preferred viewing modes (dubbed products, subtitles, or original dialogue). The second block comprises Likert-scale questions based on linguistic and extra-linguistic criteria; viewers are asked to evaluate the perceived naturalness of the synthetic voice, the perceptibility of differences between the sampled and natural voices, and the naturalness of the Turkish actor's translated speech – thus keeping the acoustic and linguistic dimensions distinct, in line with the framework outlined in Section 2. The final block explores future perspectives, asking re-

spondents if they would pay for a cinema ticket to see an AI-dubbed film, if they would watch such a product for free, and which specific genres they would be most willing to consume in this format. The questionnaire (Table 1) concludes by asking respondents whether they believe artificial dubbing will ultimately replace traditional studio dubbing.

The questionnaire items were developed by adapting the structure of the author's previous surveys (Sileo 2020, 2025) to the theoretical constructs discussed in Section 2. The final instrument draws insights from expert panels conducted alongside industry professionals and leading academics, as well as recent research on voice naturalness and the acceptance of synthetic speech (Federico *et al.* 2020; Ahmad *et al.* 2025; Spiteri Miggiani 2024, 2025; Chaume 2026). These foundations were intertwined with behavioral variables – such as willingness to pay and genre-specific acceptance – to operationalize the economic and content-related conditions of the “tolerance threshold” proposed by Spiteri Miggiani (2025). The Likert items were formulated as paired opposite statements (e.g., “utterly natural” / “utterly unnatural”) to reduce ambiguity, and the questionnaire's wording was calibrated for accessibility to non-expert respondents – the instrument was reviewed by colleagues and piloted before distribution. The stimulus material consisted of video excerpts from the AI-dubbed interviews broadcast in the 2024 edition – two video clips, broadcast on 12/10/2024, each excerpt lasting between 20 and 40 seconds². No formal psychometric validation of the instrument was undertaken; given the exploratory nature of the study, the analysis relies on descriptive statistics, and the findings are accordingly presented as tendencies within the sample rather than as inferentially generalizable results.

² Excerpt duration was intentionally capped between 20 and 40 seconds. This concise length was selected to minimize cognitive fatigue among participants and prevent potential dropouts.

Table 1. Questionnaire.

Section 1	Section 2	Section 3
Age Proficiency in Italian Education Area of expertise Job/main activity Do you watch dubbed or subtitled AV products? Do you prefer dubbing, subtitling or other modalities?	Now watch this video and select the statement you agree with. 1 The artificial voice sounds utterly natural 2 3 4 5 The artificial voice sounds utterly unnatural Based on the previous video, select the statement you agree with. 1 The difference between the original voice and the artificial one is utterly evident 2 3 4 5 The difference between the original voice and the artificial one is not evident at all Now watch this video and select the statement you agree with. 1 The contestant's speech sounds completely natural 2 3 4 5 The contestant's speech sounds completely unnatural	Based on what you have seen in the previous videos, would you go to the cinema to watch an AI-dubbed film (and thus pay for a ticket)? Would you watch an AI-dubbed product on your home TV (for free)? If you answered "yes" to the previous question, what kind of product would you watch? Do you think artificial dubbing will eventually replace traditional dubbing?

4.1. The Sample

The study presented here relies on a web-based survey administered in late 2025 via a voluntarily completed Google Form (available at <https://forms.gle/pb3H-QMfZw5YEtLtBA>). Participants were recruited through convenience sampling, supplemented by direct personal invitations, also leveraging professional networks spanning various Italian educational institutions to ensure a geographically balanced distribution. Specifically, the survey involved high school students from Vicenza, the province of Rome, and Salerno, alongside university students from the University of Milan, the University of Ferrara, Tor Vergata University of Rome, and the University of Catania, covering the country's three macro-areas (Northern, Central, and Southern Italy). Informed consent was obtained from all respondents via a mandatory confirmation checkbox prior to accessing the survey. Data collection was strictly anonymous, with no sensitive personal identifying information recorded.

Initially, the survey gathered responses from an additional age bracket (11-14) which were excluded from the final dataset. Despite efforts to use clear and accessible language in the questionnaire, the core concepts regarding AI and dubbing often proved conceptually challenging for this specific demographic, which could have compromised the validity of the data also due to insufficient sample size. Therefore, the final valid sample consisted of $N = 142$ respondents. In line with the study's focus on generational media consumption, demographic profiling was strictly focused on age, as other variables such as gender were deemed non-pertinent to the specific research questions. The respondents belong to the youth demographic (15-34 years old) and were segmented into three brackets based on Auditel standards. Specifically, the largest segment of the sample falls within the 15-19 age bracket ($n = 67$, 47.2%), followed by the 20-24 ($n = 40$, 28.2%) and, finally, by the 25-34 age groups ($n = 35$, 24.6%). This study deliberately targets younger cohorts as they are traditionally the earliest adopters of technological innovations. Consequently, the sample is composed predominantly of students. While this specific focus means the sample is not statistically representative of the whole population (see Section 7), it provides an entry point for exploring emerging reception trends.

Regarding the educational background, the largest segment holds a high school diploma ($n = 56$, 39.4%), followed by participants whose highest com-

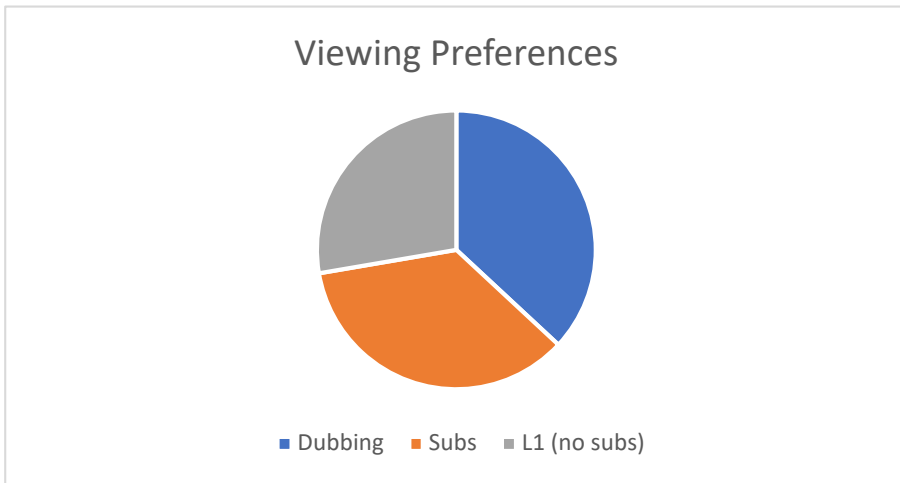


Figure 1. Viewing preferences – general average.

pleted level of education is a middle school diploma ($n = 39$, 27.5%). University-level and post-graduate education are also well-represented: 15.5% ($n = 22$) hold a bachelor's degree, while 11.9% ($n = 17$) possess advanced qualifications (master's degrees, post-graduate certificates, and PhDs). The remaining 5.6% ($n = 8$) selected the 'Other' category. Finally, in terms of employment status, most of the sample consists of students ($n = 122$, 85.9%), aligning with the targeted demographic. The remaining 14.1% ($n = 20$) are employed professionals working in various fields, including translation, psychotherapy, human resources, and communication.

To establish a baseline for media consumption, participants were asked about their frequency in watching dubbed audiovisual products. Interestingly, 14.8% of the sample ($n = 21$) stated they do not watch dubbed content at all. This subgroup is distributed across the 15-19 ($n = 6$), 20-24 ($n = 10$), and 25-34 ($n = 5$) age brackets. Conversely, heavy consumers – watching dubbed products for more than five hours a day – were extremely rare ($n = 2$, 1.4%) and exclusively belonged to the youngest demographic (15-19 years). A considerable portion of the sample ($n = 39$, 27.5%) reported moderate consumption, watching dubbed products only “a few hours a month”, comprising 24 respondents from the 15-19

group, 4 from the 20-24 group, and 11 from the 25-34 group. This variance in habitual viewing patterns provides a crucial contextual background for analyzing their attitudes towards AI-generated dubbing.

When examining the participants' primary viewing preferences, the overall sample is relatively well-distributed across different modalities. Figure 1 shows the general average, with respondents preferring dubbed content closely followed by those who opt for subtitles, whereas a substantial amount of interviewees favor watching content in its original language without subtitles.

However, breaking down these preferences by age group reveals generational divergences. The youngest demographic (15-19) shows a strong preference for dubbing ($n = 32$), with the remainder equally split between original language without subtitles ($n = 15$) and subtitled content ($n = 15$). Conversely, the 20-24 age bracket strongly favors subtitles ($n = 20$), compared to original language ($n = 11$) and dubbing ($n = 7$). A similar trend is observed in the oldest cohort (25-34), where subtitles ($n = 11$) and original language without subtitles ($n = 10$) slightly outpace the preference for dubbing ($n = 9$).

5. Results

5.1. Perception of the Artificial Voice's Naturalness

Moving to the core experimental phase of the study, participants were asked to evaluate the perceived naturalness of the artificial voice in the AI-generated dubbing sample on a 5-point Likert scale, where 1 indicates an "absolutely natural" perception and 5 an "absolutely unnatural" one. An initial analysis of the mean scores reveals a general tendency to perceive the AI voice as relatively natural, with all age brackets scoring below the neutrality threshold of 3.0. Specifically, the 20-24 age group proved to be the most receptive to the AI's naturalness, recording a mean score of 2.55. The youngest (15-19) and the oldest (25-34) cohorts showed very similar mean scores of 2.70 and 2.71, respectively.

To better understand the distribution of these attitudes, the responses were aggregated into perceived naturalness and perceived unnaturalness, isolating the neutral responses. The data reveals a surprising level of acceptance regarding the

voice's acoustic quality. Among the 15-19 demographic, respondents evaluated the voice as mainly natural. This trend is even more pronounced in the older cohorts, as displayed in Figure 2.

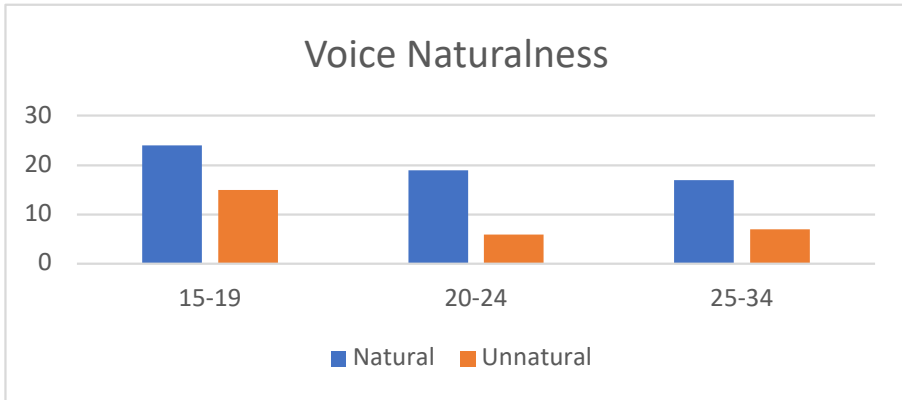


Figure 2. Voice naturalness as perceived across age groups.

A deeper qualitative look at the extreme negative responses reveals that the highest possible score for artificiality was almost entirely absent across the entire sample. It was selected only twice within the 15-19 group, and never by the 20-24 cohort, suggesting that the AI-generated dubbing avoided the extreme robotic perception typically associated with traditional text-to-speech technologies.

5.2. Detectability of the Sampled Voice

Participants were subsequently asked to evaluate the detectability of the AI manipulated voice when compared directly to the human one. This was measured on a 5-point Likert scale, where 1 indicated that the difference between the original and artificial voice was “absolutely evident,” and 5 indicated that it was “not evident at all.”

The mean scores reveal a clear generational progression in the ability to detect the AI-generated voice. The oldest cohort (25-34) demonstrated the highest level of discernment, with a mean score of 2.37, leaning towards the recognition

of the artificial nature of the sample. This detectability slightly decreases in the 20-24 age group (2.90), while the youngest demographic (15-19) positions itself almost exactly on the threshold of neutrality (3.07), suggesting a lower sensitivity to the difference between the natural and the synthetic voice. To isolate the extremes and avoid the central tendency bias, the data was grouped into “evident difference” and “imperceptible difference.” The analytical breakdown is consistent with the age-related trend (see Figure 3).

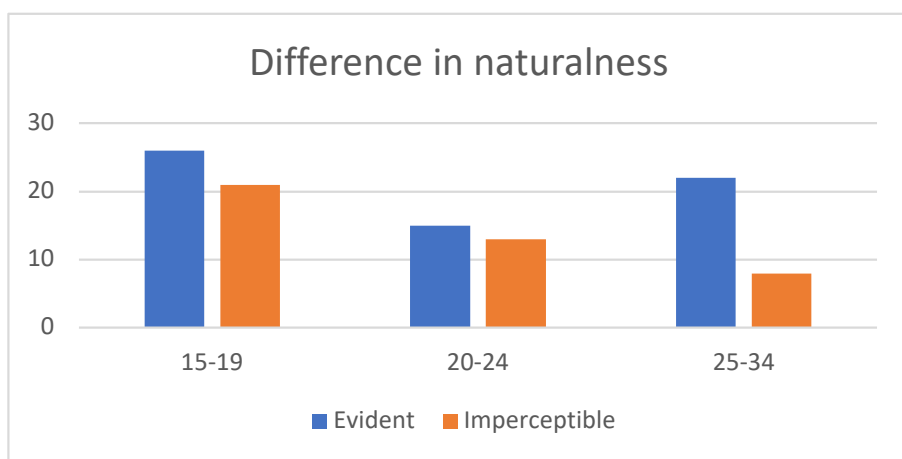


Figure 3. Artificial voice detectability across age groups.

Whereas opinions were highly polarized in the intermediate 20-24 group, and slightly less so in the 15-19 bracket, a significant amount of 25-34 respondents easily distinguished the AI voice from the original, with only a small portion who found the difference hard to detect.

5.3. Naturalness of Linguistic Choices and Phrasing

The final parameter of the experimental block shifted the focus from the acoustic quality of the voice to the linguistic dimension of the AI dubbing. Participants were asked to evaluate the naturalness of the contestant’s specific linguistic

choices, as generated by the AI. Responses were recorded on a 5-point Likert scale, where 1 denoted “totally natural expressions” and 5 indicated “totally unnatural expressions.”

The mean scores indicate a general shift towards skepticism when evaluating the AI’s linguistic output, with all age groups scoring above the neutrality threshold of 3.0. This presents a noteworthy contrast with the earlier acoustic evaluation: while the AI voice may sound superficially human, the generated script and target-language phrasing are perceived as comparatively unnatural. The 20-24 demographic emerged once again as the most critical of the AI’s linguistic choices, recording the highest mean of 3.27. The youngest group (15-19) and the oldest cohort (25-34) exhibited similar stances, leaning slightly towards artificiality, with mean scores of 3.07 and 3.02, respectively.

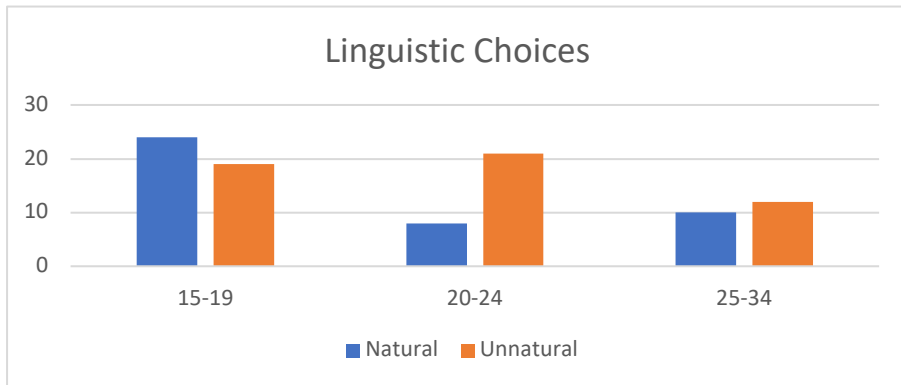


Figure 4. Naturalness of linguistic choices across age groups.

By aggregating the responses to isolate clear trends, the data was grouped into “natural linguistic choices” and “unnatural linguistic choices.” The AI’s phrasing was found unnatural mainly in the 20-24 group, where the rejection was stark, whereas the other age brackets demonstrated a moderate division. Interestingly, a qualitative look at this oldest cohort reveals that despite leaning towards unnaturalness, not a single participant selected the extreme negative score (5, “totally unnatural”).

5.4. Willingness to Pay

Transitioning from perceptual evaluation to behavioral intention, section 3 of the survey investigated the participants' "willingness to pay" for AI-generated audio-visual products. When asked if they would go to the cinema and purchase a ticket to watch an AI-dubbed film based on the samples provided, the response was overwhelmingly negative across the board, signaling a strong market resistance.

The data reveals that the 20-24 demographic is the most resolutely opposed: a striking 77.5% ($n = 31$) stated they would not pay for a cinema ticket, with only 12.5% ($n = 5$) willing to do so, and a marginal 10.0% ($n = 4$) remaining undecided. This firm rejection correlates perfectly with their highly critical evaluation of the AI's linguistic and expressive unnaturalness in the previous experimental block. The other two cohorts display a slightly softer, yet still predominantly negative, stance. Among the youngest group (15-19), despite their proven susceptibility to the auditory illusion of the AI, a solid 58.2% ($n = 39$) refused the idea of paying for a theatrical release, with only 14.9% ($n = 10$) accepting it and 26.9% ($n = 18$) who would not know. A virtually identical trend is mirrored in the oldest demographic (25-34), where 60.0% ($n = 21$) rejected the paid cinema experience, 14.3% ($n = 5$) accepted it, and 25.7% ($n = 9$) expressed uncertainty. Ultimately, these figures suggest that while AI dubbing might pass casual acoustic scrutiny, it currently falls severely short of the quality threshold expected by audiences for paid cinematic consumption.

5.5. Acceptance of Free Home-Viewing

To verify whether the widespread rejection of the AI dubbing was driven by ideological resistance or by an economic motive, participants were subsequently asked if they would watch an AI-dubbed product for free on their home television. The removal of the premium cost and the shift to a more casual, domestic viewing context produced a dramatic change in audience acceptance.

The most striking reversal occurred within the youngest demographic (15-19), where acceptance became the relative majority. A significant 46.3% ($n = 31$) stated they would watch the free AI-dubbed product, compared to 35.8% ($n = 24$) who

still refused, and 17.9% ($n = 12$) who were unsure. The oldest cohort (25-34) exhibited a similar paradigm shift, moving from a 60.0% rejection rate for cinema to a slight positive margin for home viewing, with 37.1% ($n = 13$) answering “yes,” 34.3% ($n = 12$) answering “no,” and 28.6% ($n = 10$) expressing uncertainty.

Interestingly, the 20-24 age group confirmed its status as the most critical and resistant cohort of the study. Even when presented with a free, home-viewing scenario, it remained the only demographic where the outright refusal (35.0%, $n = 14$) still marginally outweighed the acceptance (32.5%, $n = 13$), alongside a high level of indecision (32.5%, $n = 13$). Overall, these findings strongly suggest that the current threshold of AI dubbing quality is deemed acceptable for low-engagement, free-of-charge consumption, but it lacks the premium quality required for paid theatrical distribution.

5.6. *Preferred Genres for AI Dubbing*

To further contextualize the acceptance of free AI-dubbed content, respondents who were open to the idea (including those who were previously undecided) were asked to indicate which specific audiovisual genres they would be willing to watch. The distribution of preferences, allowing for multiple selections, reveals a clear hierarchy of acceptability based on the emotional and performative demands of the format.

Across all age groups, non-fiction unscripted formats emerged as the safest avenues for AI dubbing. “Docu-reality” was the most frequently selected genre overall, gathering 20 preferences in the 15-19 bracket, 10 in the 20-24 bracket, and 8 among the 25-34 group. Documentaries and reality shows also received consistent support. This strongly suggests that audiences intuitively recognize the current limitations of AI in delivering deep emotional performances, relegating the technology to factual or voiceover-driven content where informational delivery supersedes complex acting.

However, a notable generational divide emerges regarding scripted content like traditional films. The youngest demographic (15-19) showed a surprising openness to AI-dubbed movies, which gathered 18 selections, making it their second most preferred option. In stark contrast, older viewers severely restricted the scope of

AI: films received only 6 preferences in the 20-24 group and a mere 4 in the 25-34 group, confirming their demand for human authenticity in cinematic storytelling.

Finally, the open-ended responses collected in this section provided crucial qualitative insights. Within the highly critical 20-24 cohort, the ethical dimension of AI was explicitly raised, with a participant strongly objecting to the technology due to the risk of human job displacement in the dubbing industry. Similarly, the 25-34 bracket introduced strict conditional boundaries, with participants stating they would accept AI only for “small fragments” or under the strict condition that “it is done well.” These remarks highlight that, for young adult audiences, technological adoption is not merely a matter of acoustic quality but is heavily filtered through ethical concerns and stringent quality-control standards.

5.7. The Future of Dubbing

The survey concluded by asking participants to project current technological trends into the future. The responses suggest that, despite rapid advancements in voice generation, the respondents in this sample do not currently envisage a full replacement of human performance.

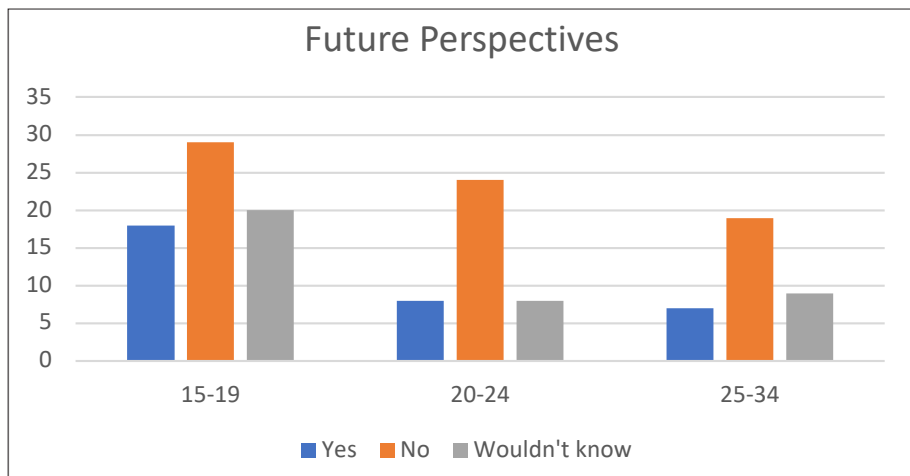


Figure 5. Future perspectives for the local AVT industry.

Across all demographics, the prevailing sentiment is that AI will not entirely replace traditional dubbing. Figure 5 shows how the 20-24 age group – which consistently proved to be the most critical of the AI’s expressive capabilities throughout the study – exhibited the strongest conviction, mainly rejecting the idea of total replacement. The 25-34 cohort mirrored this stance, while the youngest demographic (15-19) displayed a slightly more open, albeit still predominantly skeptical, attitude towards a fully automated future.

Ultimately, these findings synthesize the core narrative of the reception study: while AI dubbing is gaining ground as a viable, cost-effective alternative for secondary or unscripted content, the target audience remains critically aware of its performative limitations. The public surveyed here recognizes AI as a supplementary tool in the audiovisual landscape, rather than a definitive substitute for the artistry and emotional resonance of traditional human dubbing.

6. Discussion

The narrative emerging from the data reveals a complex, highly stratified audience reception to AI dubbing. The findings point to a dichotomy in AI performance: while current technologies successfully mimic the acoustic parameters of human naturalness, the technology falls short when it comes to linguistic choices. The audience, particularly the highly critical 20-24 demographic, appears to distinguish the physical sound of the voice from the linguistic competence required for professional dubbing. In terms of the framework outlined in Section 2, respondents proved more tolerant of the acoustic layer of the “triple duplicity” than of its linguistic layer: the synthetic voice relatively passed unnoticed compared to the AI-generated translation.

A defining outcome of this study is the stark generational divide regarding the detectability and acceptance of artificial voices. As previously hypothesized, younger digital natives (the 15-19 cohort) – who are arguably more accustomed to rapid, fragmented media consumption and the pervasive use of synthetic voices on TikTok and Instagram – exhibit a significantly lower sensitivity to the discrepancy between human and AI-generated voices. Their daily media diet has seemingly normalized artificial audio, making them more susceptible to the au-

ditory illusion and more permissive toward free, AI-dubbed content. In contrast, the young adult cohorts (20-34), who grew up with traditional, high-quality human dubbing, maintain a more refined auditory baseline and stronger ethical reservations.

Finally, the study clearly demarcates the current commercial boundaries of AI in audiovisual translation. The unanimous rejection of the AI-dubbed cinematic experience, contrasted with a moderate acceptance of free home-viewing for unscripted formats (e.g., docu-realities, documentaries), proves that this audience views AI not as an artistic equivalent to human talent, but as a functional, low-cost tool. Ultimately, these viewers fiercely protect the emotional authenticity of cinematic storytelling, relegating AI to factual delivery where deep emotional resonance is not the primary requirement. It should be stressed, however, that the patterns identified here may represent broader tendencies that warrant further testing through large-scale future research.

7. Conclusion

This study set out to explore the perceptual and behavioral dynamics surrounding AI-generated dubbing in Italy. By evaluating the acoustic, linguistic, and economic parameters of reception, the findings attempted to provide empirical answers to the research questions and hypotheses that guided this investigation.

Addressing RQ1 and RQ4 – concerning the genuine embrace of the technology and the relevance of economic motivations –, the data strongly validates RH4, while adding crucial nuance to RH1. Artificial dubbing is indeed experiencing a slow, conditional acceptance within the local audiovisual landscape, but this acceptance is intrinsically tied to the economic barrier of premium consumption. The overwhelming rejection of paying for an AI-dubbed cinema ticket, sharply contrasted by a sudden openness to free home-viewing, confirms that economic motivations dictate audience tolerance. AI is currently perceived as a low-cost, functional alternative to some artistic product worthy of financial investment.

Regarding the role of audiovisual genres (RQ3), the hypothesis that acceptance would extend to scripted content such as films (RH3) was largely refuted

by the older cohorts (20-34) but surprisingly supported by the youngest demographic. While the consensus generally relegates AI to non-fiction and unscripted formats (e.g., docu-realities and documentaries) where emotional performance is secondary, the 15-19 age group demonstrated a unique willingness to consume AI-dubbed traditional films. This highlights a fracturing audience landscape where digital natives' media diets influence their format tolerance.

Finally, investigating the ultimate trajectory of the industry (RQ2) and the potential replacement of traditional studio dubbing initially among younger audiences (RH2), the findings offer a reassuring outlook for human professionals. While the 15-19 cohort proved to be the most susceptible to the auditory illusion and the most permissive in terms of consumption, they did not predominantly support the idea of total technological replacement. The highly critical young adult demographic (20-24) firmly led the defense of human artistry. Therefore, RH2 is refuted in its extreme premise: while younger audiences act as early adopters, the overarching consensus remains that AI will serve as a supplementary tool for secondary content, rather than an eventual substitute for the emotional resonance and linguistic authenticity of traditional, human-led dubbing.

By way of conclusion, some limitations of the present study must be acknowledged. First, the investigation is based on a single AI-dubbed sample from the 2024 edition of *Ballando con le stelle*. As a non-scripted format, this specific program may have inherently shaped audience expectations and responses in ways that differ significantly from scripted fiction. Consequently, the findings cannot be generalized beyond this specific televised context without due caution. Second, the sample, recruited among younger audiences, cannot be considered statistically representative of the whole Italian viewing public; self-selection bias cannot be excluded. Third, the analysis relies on descriptive statistics, so the generational differences identified should be read as tendencies rather than statistically established effects.

Future research should test these tendencies on a wider range of AI-dubbed genres – including scripted fiction – with larger and more diverse samples, complementing quantitative instruments with qualitative methods (e.g., interviews or focus groups) and inferential analysis. Despite these limitations, the study offers a first empirical attempt at mapping Italian audience reception of AI-driven dubbing at a moment when the phenomenon is only beginning to reach main-

stream broadcasting, and it is hoped that it may serve as a baseline for the reception studies that this rapidly evolving scenario will undoubtedly demand.

References

- Ahmad, Maysa, Ahmad S. Haider, and Hadeel Saed. 2025. "Assessing AI driven dubbing websites: Reactions of Arabic native speakers to AI dubbed English videos in Arabic." *Research Journal in Advanced Humanities* 6(1).
- AIDAC 2023. NUOVO CCNL DOPPIAGGIO - AIDAC. (25/2/2026).
- Bryant, Miranda. 2025. Denmark to tackle deepfakes by giving people copyright to their own features | Deepfake | The Guardian (10/7/2026).
- Bucaria Chiara and Delia Chiaro. 2007. "End-user Perception of Screen Translation: The Case of Italian Dubbing." *TradTerm* 13 (2007): 91-118.
- Chaume, Frederic. 2026. *Dubbing. Concepts and Practices*. 2nd edition. London: Routledge.
- Di Francesco, Valentina and Angela Sileo. 2025. "Simil Sync/Linear Sync: Past, Present, and Future." *Journal of Audiovisual Translation* 8(2): 1-22.
- Di Giovanni, Elena and Yves Gambier, eds. 2018. *Reception Studies and Audiovisual Translation*. Amsterdam/Philadelphia: John Benjamins.
- Federico, Marcello, Robert Enyedi, Roberto Barra-Chicote, Ritwick Giri, Umut Isik, Arvinth Krishnaswamy, and Hassan Sawaf. 2020. "From speech-to-speech translation to automatic dubbing." In *Proceedings of the 17th International Conference on Spoken Language Translation*, 257-264. Ed. by Marcello Federico, Alex Waibel, Kevin Knight, Satoshi Nakamura, Hermann Ney, Jan Niehues, Sebastian Stüker, Dekai Wu, Joseph Mariani, Francois Yvon. Association for Computational Linguistics.
- Gambier, Yves. 2018. "Translation Studies, Audiovisual Translation and Reception." In *Reception Studies and Audiovisual Translation*. Ed. by Elena di Giovanni and Yves Gambier, 43-66. Amsterdam/Philadelphia: John Benjamins.
- Goodfellow, Melanie. 2025. AI-Generated Actress Could Be First to Be Signed by Talent Agent. Deadline. (10/7/2026).
- Kovačić, Irena. 1995. "Reception of subtitles. The non-existent Ideal viewer." *Translatio (Nouvelles de la FIT/FIT Newsletter)* 14 (3-4): 376-383.
- Monaco, Antonio. 2025. Voce clonata con AI, il figlio del doppiatore Claudio Capone diffida formalmente la Rai. HDBlog. (25/2/2026).
- NDTV 2025. "I Don't Exist": British News Channel Anchor Reveals She Is AI-Generated. NDTV. (25/2/2026).

- Pavesi, Maria. 2008. "Spoken Language in Film Dubbing. Target Language Norms, Interference and Translation Routines." In *Between Text and Image. Updating Research in Screen Translation*. Ed. by Delia Chiaro, Christine Heiss, Chiara Bucaria, 79-99. Amsterdam-Philadelphia: J. Benjamins, 79-99.
- Raffaelli, Sergio. 1983. *Le parole proibite. Purismo di stato e regolamentazione della pubblicità italiana (1812-1945)*. Bologna: Il Mulino.
- Raffaelli, Sergio. 1991. *La lingua filmata. Didascalie e dialoghi nel cinema italiano*. Firenze: Le Lettere.
- Sileo, Angela. 2018a. "Il simil sync o semi-sinc nel panorama italiano: lo slittamento e offuscamento dei confini previsti dal CCNL." *Altre Modernità*: 258-266.
- Sileo, Angela. 2018b. "Doppiaggese": *Verso la costruzione di un metodo*. Collana Ricerca Continua, Vol. II. Rome: UniversItalia.
- Sileo, Angela. 2020. "Dubbing or Simil Sync? A Study on Reception in Italy." In *Reception Studies and Adaptation: A Focus on Italy*. Ed. by Giulia Magazzù, Valentina Rossi, and Angela Sileo, 174-190. Cambridge: Cambridge Scholars Publisher.
- Sileo, Angela. 2024. "Adapting Oral Markers in Mockumentary Dubbing: A Case Study of *Modern Family* (2009-2020)." *Lingue e Linguaggi* 67: 91-105.
- Sileo, Angela. 2025. "Alterity in AVT: Flipping the Coin of (In)Visibility." In *Adapting alterity in anglophone scenarios*. Ed. by Angela Sileo and Giulia Magazzù, 41-59. Wilmington (DE): Vernon Press.
- Spiteri Miggiani, Giselle. 2024. "Quality assessment tools for studio and AI-generated dubs and voiceovers." *Parallèles* 36(2): 50-70.
- Spiteri Miggiani, Giselle. 2025. Rethinking creativity in dubbing. Potential impact of AI dubbing technologies on creative practices, roles and viewer perceptions. *Translation Spaces*. John Benjamins.
- Toury, Gideon. 1995. *Descriptive Translation Studies – And Beyond*. Amsterdam: John Benjamins.
- Tuveri, Matteo. 2025. Pubblicità Arredissima Gerry Scotti, recensione, descrizione. Marida Caterini. (10/7/2026).