Journal of Audiovisual Translation

Easy Audios: From Readability to Listenability

Machuca, María J. ✉
Universitat Autònoma de Barcelona

Matamala, Anna ✉
Universitat Autònoma de Barcelona

Citation: Machuca Ayuso, M. J., & Matamala, A. Easy Audios: from Readability to Listenability. Journal of Audiovisual Translation, 8(1), 1–19. https://doi.org/10.47476/jat.v8i1.2025.345

Editor(s): N. Reviers

Received: October 17, 2024

Published: April 11, 2025

Abstract

This article provides a foundational exploration of key concepts related to the development of "easy audios", an emerging accessibility service based on Easy Language (EL). EL encompasses a set of guidelines for simplifying language to enhance accessibility for individuals with reading comprehension difficulties, applicable across various formats, including written, oral, and multimodal content. The article examines the transition from easy written texts to easy audios, focusing on readability and listenability, crucial concepts in both foreign language learning and audiovisual accessibility. The study highlights existing practices in easy subtitles and audio descriptions and discusses innovative hybrid services such as easy audios, which combine EL and audiovisual translation principles. Furthermore, the research explores how easy audios could be developed for improved comprehension, particularly for those who struggle with audiovisual content. Conclusions suggest that essential features for listenability in easy audios have been identified, highlighting the importance of clear linguistic structures, appropriate sound mixing, and careful prosody. However, the article emphasises the need for further research to fully understand and refine these features, particularly as they relate to the development of easy audio services in various contexts.

Key words: easy language, listenability, comprehensibility, audiovisual translation, accessibility, pronunciation, prosody.

✉ mariajesus.machuca@uab.cat, https://orcid.org/0000-0003-4377-288X
✉ anna.matamala@uab.cat, https://orcid.org/0000-0002-1607-9011

Introduction

According to the international standard ISO/IEC 23859, Easy Language (EL) can be defined as "a language variety in which a set of recommendations regarding wording, structure, design and evaluation are applied to make information accessible to people with reading comprehension difficulties for any reason". A note to the definition clarifies that Easy Language is often referred to as "Easy-to-Read", although the document prefers the term "Easy Language" as it can be applied not only to written but also to oral or multimodal content.

EL enhances comprehensibility and may be useful for persons who struggle to understand written and spoken language. This difficulty may be due to the context in which the interaction takes place, but also to users' different abilities, previous knowledge, or interest in the topic. EL is expected to benefit persons with intellectual disabilities, low literacy levels, language learners, or persons in stressful situations, to name a few examples.

EL recommendations have generally mimicked existing practices, which have focused on written content. Nevertheless, references are also made to other formats, such as video or audio information (Inclusion Europe, 2009), which are often seen as alternative formats (Accessible Information Working Group, 2011). Lindholm and Vanhatalo (2021) co-edited an overview of EL in Europe which shows the predominance of written content in EL (public administration, literature, informative texts, media, etc.) over audio or audiovisual content, which nevertheless has some presence on the radio, television, and web. Matamala (2023) provides an overview of such practices and lists, for example, the ORF Austrian news, the YLE Easy Finnish TV news, the SVT Easy Swedish news, or the Latvian Easy radio news.

Merging EL and audiovisual accessibility and translation services as well as new hybrid services, have been suggested (Bernabé-Caro & Orero, 2019), including easy audio descriptions, easy subtitles, or easy audio subtitles. These access services have been the object of limited research, which has focused on easy subtitles (Bernabé-Caro, 2020; Oncins et al., 2020) and easy audio descriptions (Bernabé-Caro & Orero, 2020; Arias-Badia & Matamala, 2020, 2023), following the EASIT research project (Matamala, 2022). Reference to easy interpreting is also found, with Schulz et al. (2020) describing how interpreters become "clarifiers" and add explanations to guarantee comprehension for persons with cognitive disabilities. Similarly, Nahón Guillén (2020) describes an experience termed "simultaneous simplification", which Yalon-Chamovit & Avidan-Ziv (2016) define as "the structured process of processing and editing information so as to make it clear, simple and readily understood by persons with cognitive disabilities" (p. 1).

A related service mentioned in the literature is "easy audios" or "audio explanations". These easy audios are inserted in the silent gaps—similarly to audio description—to enhance comprehension by explaining difficult words or what is happening. They are created using EL recommendations and validated by users with cognitive disabilities. A closely related concept is that of "audio description

for persons with intellectual disability" in Argentina, featuring an audio track that explains what is being said and what is happening in a video in simple language.

Building on these experiences, Matamala (2023) explores the development of easy audio access services alongside existing audiovisual modes, highlighting easy audios as a service worth studying. These audios target individuals who can access audio and visuals but struggle to fully understand the content. An off-screen narrator could be used, primarily in silent gaps, but also overlapping where necessary, to enhance understanding. Key elements would include linguistic features, sound mix, and prosody. Easy audios could be used, for instance, to make theatre plays or television programmes easier to understand by clarifying what is going on. However, a museum could also offer an easy audio guide. The reality is that the concept of easy audio is not fully established, lacks an agreed definition, and merits further research.

Easy audios are the object of research of the WEL project (2023–2026), funded by the Spanish Ministry of Science, Innovation and Universities and co-funded by the European Union (PID2022-137058NB-I00). WEL aims to research the specificities of EL when moving from the written to the oral by: a) mapping current oral EL practices in terms of creation processes, users, and features; b) evaluating selected oral EL features in terms of comprehension, preferences, and acceptability; c) researching how easy audios could be developed considering existing access services and using video games and cultural leisure visits as case studies. However, the first step is to establish the conceptual basis.

This article aims to provide an overview of key notions such as legibility, readability, intelligibility or listenability as the basis for future research on easy audios. First, the article focuses on the concepts of readability and listenability, exploring how they can be defined in the context of written and spoken language. The specificities when moving from easy written texts to easy audios are explored. A section is devoted to how these concepts have been researched in foreign language learning, as this is one of the areas where much work is found. Then, the article deals with the concept of listenability in audiovisual accessibility and presents some recommendations on easy spoken language. The paper concludes by suggesting some features that could be related to listenability, whilst acknowledging the need for further investigations.

1. From Readability to Listenability

When moving from research on the written to the oral, some key concepts need further investigation and comparison. Comprehensibility is a central aspect in both oral and written language as it aims to identify the parameters that make texts easy or hard to understand (Wolfer, 2015). As described by Friedrich and Heise (2022, p. 2), comprehensibility was linked in previous models to inherent features of the texts, but an interactionist view of comprehensibility, considering the reader-text interaction, seems to be more aligned with current approaches. Comprehension, on the other hand, refers to "the process of understanding a text by building a mental representation" (Wolfer, 2015, p. 34).

In relation to written texts, two concepts emerge: legibility and readability. Whereas legibility concerns the "visual perception of typeface and layout" (DuBay, 2007, p. 4), readability is "the ease of reading created by the choice of content, style, design, and organization that fit the prior knowledge, reading skill, interest, and motivation of the audience" (DuBay, 2007, p. 6). Many applications test or predict the readability of a text, and different metrics are used to provide a readability score. Measures like the Flesch-Kincaid Grade Level (Flesch, 1948), the Gunning Fog Index (Gunning, 1952), syllable count, word count, sentence and paragraph length, and spelling/grammar errors are considered. They provide readability scores and information about quality issues and suggest how to make a text more readable. Although there are different metrics to predict the readability of a text, connecting readability and comprehension is still an issue.

In relation to the spoken word, the concepts of intelligibility and listenability are found. Although different definitions exist (Cruz, 2007), in our paper intelligibility refers to how easy and accurately a listener can recognise a sound. In terms of spoken language clarity, it seems there is a point beyond which pronunciation becomes critical, and individuals who do not meet this standard are likely to encounter difficulties in communication (Karas, 2020). Baese-Berk et al. (2023) view intelligibility measures as direct measures of speech perception, i.e. "correctly identifying the linguistic items (phonemes, words, sentences) produced by the speaker" (p. 68). Factors affecting speech perception include background noise, talker variation and familiarity (including accents and speech disorders), listener population (e.g., hearing loss, age), sentence predictability, lexical frequency, and individual differences. As for listenability, it concerns the linguistic features that make a piece of spoken language more easily understood by both native speakers and language learners. As pointed out by Messerklinger (2006), listenability is influenced not only by factors shared with readability, such as linguistic complexity and vocabulary, but also by phonetic, acoustic, auditory and articulatory factors, and by the listener's background knowledge. Listenability for a learner may vary depending on their level of language proficiency, familiarity with the language, and ability to adapt to different accents and speech styles. For this reason, it is difficult to establish metrics for listenability, which is often assessed through comprehension tests.

Generally speaking, readability in written text corresponds to listenability in oral productions. Still, the extent to which readability and listenability correlate is still not clear (Denbow, 1975; Eastwood & Snook, 2012; Kotani et al., 2014). As expressed by Glenn et al. (1995, p. 46), many authors have used the term "listenability" in their articles and state that "all seek something about the clarity, ease, and agreeableness of aural stimuli as perceived by listeners". Early work on listenability focused on readability formulas, such as the work of Denbow (1975), Harwood (1955), or Horowitz and Samuels (1985). Rubin (1993) took a step further and thought that readability formulas are not suitable in this context. Rubin (2012, p. 178) states that, despite some reading research that considers listening as "reading by ear" (Sticht & James, 1985), the listening process presents both challenges and advantages in comparison to reading. Rubin mentions that "listeners can take advantage of prosodic cues to help clarify the informational structure of discourse units" but at the same time are "generally at the mercy of the speaker's rate production". According to the same author, a listenable discourse is "characterized by linguistic and rhetorical structures that ease the particular cognitive burdens

listeners face" (2012, p. 178). Rubin proposes a Listenability Scale Guide, a list of suggestions rather than requirements, with many commonalities with EL recommendations. The Listenability Style Guide is divided into four domains:

- Sentence Structure: Use coordination over subordination, keep clauses moderate in length, and place subordinate clauses at the end. Prefer verbs over dense noun forms.
- Vocabulary: Use personal pronouns, everyday words, and low lexical diversity. Repeat key nouns and verbs.
- Conversation Features: Use contractions, start sentences with conjunctions, and ask questions to engage. Include tag questions, simple idioms, and address the listener by name.
- Listener Consideration: Summarise internally, provide advance organisers, signal transitions, and announce important topics. Be redundant and use stories and vivid analogies.

However, some key features of spoken language, such as speech rate, pausing, or pitch, are neglected.

2. Readability and Listenability in Teaching a Foreign Language

Although our focus is on easy audios in audiovisual translation (AVT) and accessibility, there is substantial research on the topic in foreign language learning, hence our interest in providing a summary of how readability and listenability are approached in this area.

Foreign language teachers are concerned about the texts' readability according to their students' proficiency. Readability considerations, such as appropriate vocabulary and sentence structure, can facilitate understanding, making the learning process more efficient and enjoyable. However, if texts are too complex or include unfamiliar vocabulary, learners may struggle to understand the content, delaying their language acquisition process. In language acquisition, differences between first language (L1) acquisition and second language (L2) learning are evident from the outset (Beinborn et al., 2014; Xia et al., 2016). L1 is learned informally through everyday communication, while L2 requires formal instruction and deliberate tactics. L2 readability assessments often fail, as they are based on native speaker data, not L2 learners because there is not a significantly sized corpus for learners (Xia et al., 2016).

Assessing textbook readability is vital for language teaching. Handayani and Wirza (2020) employed the Flesch-Kincaid Readability Formula to analyse language content and readability levels in a textbook. Their findings indicated that only five texts were at an appropriate readability level, while six were below the expected grade and age level, and two were above. These results emphasise the significance of selecting texts that align with students' cognitive and linguistic capacities, striking a balance between challenge and comprehension for effective learning. Automatic readability models

for L2 use algorithms to measure text complexity, helping educators choose suitable materials. However, validating these models requires learner-specific corpora, which can be challenging to obtain (Xia et al., 2016).

Listenability in a foreign language refers to how easily learners can understand spoken language, affecting their comprehension of conversations, lectures, and audio materials. Key considerations for listenability in L2 include clear pronunciation since learners improve their pronunciation from listening. Kotani et al. (2014) suggest that future research should focus on creating an automatic system to assess pronunciation and subsequently examine its application within language teaching classrooms, avoiding the subjective point of view of the teachers.

Speech rate is another key point for comprehension. Speech that is too fast can be hard to follow, while speech that is too slow may reduce engagement. An optimal pace helps learners process information effectively. Chiu and Chen (2023) conducted an experiment to determine whether slowing down the speed of speech resulted in more easily comprehensible input. Speech delivery, initially set at 116 words per minute (wpm) according to listening comprehension tests of the General English Proficiency Test, was adjusted to slower rates of 98 wpm and 58 wpm for two separate test groups. The results revealed a significant positive effect on listening comprehension at the slightly slower rate of 98 wpm, indicating that 116 wpm might be a speech rate too fast for young EFL listeners. Slower speech was found to improve comprehension and boost learners' self-confidence.

Like readability for texts, listenability focuses on using suitable vocabulary and sentence structure to enhance listeners' comprehension. Aligning speech with learners' proficiency and gradually increasing complexity improves understanding. Loukina et al. (2016) found that an automatic system relying on text complexity features could predict item difficulty as accurately as, if not better than, human judgments. They found that vocabulary features like word frequency and specificity were key to item difficulty. However, focusing only on these features weakened the system's predictive performance compared to using a broader range of factors.

3. Listenability in Audiovisual Translation and Accessibility

AVT and accessibility deal with access services that cater for linguistic and sensorial accessibility, such as dubbing, voice-over, oral language and sign language interpreting, audio description, audio subtitling, and subtitling (Matamala, 2019). More recently, cognitive accessibility has also been put in the spotlight, and the hybridisation of access services with easy-to-understand languages and the study of new access services such as easy audios have been put forward.

In the field of audiovisual accessibility, the concept of listenability is referred to in some publications on audio description. Perego and Taylor (2022) state that "[g]ood vocal delivery is in turn crucial, especially to ensure the comprehension and listenability of end-users" and add: "Listenability is a measure of the ease or pleasure in listening to something, such as music, or even an oral text. Audio descriptions should always display this quality in order to be enjoyed and understood as widely as

possible". The authors consider that it is "difficult to list all the features a listenable AD should contain", but they refer to: a) slow pace and volume; b) clear and engaging tone of voice; c) removal of all interfering sounds. They consider the voice talent should "use an oral-based style, not read too fast, convey engagement, adapting if necessary to the style and genre of the source text". They also mention additional elements such as the "selection of an appropriate voice for aural reproduction", "the calibration of the volume", "the quality of the recording", and "the adjustment of the rate of delivery" (p. 43).

As part of the EASIT project, Navetta (2021) interviewed four professional describers in Germany, the United Kingdom, and the United States, as well as Prof. Chris Taylor on the concept of listenability. When asked if they had ever heard the term, the German and British professionals were unfamiliar with the term in relation to AD but could infer its meaning. The American describer recognised it, linking it to tone, pace, and voice quality. The academic noted it had become a "buzzword in audiovisual circles" alongside readability and comprehensibility. Key aspects for making AD easier to listen to included sentence structure, vocabulary choice, pace, sound mix, natural voices, appropriate prosody and tone, and adjustable settings for users.

The concept of listenability is the object of a dedicated section in a chapter on AD for the arts by Perego (2023, p. 25), where she states that the "listener-friendliness of a message delivered orally is determined by several factors, and it is called listenability", while acknowledging this term is "still not yet used systematically in the AD context" and refers to the already mentioned interview by Navetta. The Italian researcher considers that listenability is "a decisive listener-dependent feature of all aural media" which can "contribute to facilitating, or even accelerating, the active process of listening (as opposed to the passive process of hearing), which involves the recognition of a sound-meaning correspondence" (p. 62). Perego acknowledges the lack of a measure for listenability and considers that listenability "primarily relates to the prioritisation of textual, linguistic, rhetorical (Rubin, 2012, p. 177) and sound-related factors (Weaver, 1972)", with the role of the talker being crucial. The key features that Perego mentions based on a literature search are a simple oral-based language style, good text organisation, and adequate listening-engaging strategies. Perego provides an overview of specific strategies that can enhance listenability, including the recommendation that "a listenable text should be read out expressively, clearly, and fluently. This encompasses the accurate reading of difficult or technical words, a tone of voice congruent to the content, a clear prosodic segmentation of difficult words and complex texts" (2023, p. 64), and she adds: "changing pace (speeding up and slowing down), volume (saying some words louder than others), and intonation can help the reader emphasise parts of the text, thus facilitating information processing" (2023, p. 64). Perego considers that the speech should not be too fast or too slow, and pauses and sound mix are critical, but she also acknowledges that personalisation on delivery rate or speech clarity can help users (Orero, 2022).

In a more recent article on podcasts, Perego (2024, p. 458) uses Rubin's definition of listenability and describes it as a "decisive feature of all aural media which measures the 'quality of discourse that

eases the cognitive burden that aural processing imposes' (Rubin, 2012, p. 176)". Some of the features Perego (2024) identifies as facilitating towards this listenability are

> the use of an oral-based language style, signpost language guiding listeners through the audio text, a coherent text structure, a wise choice of (clear yet engaging) words, and the skilled use of most voice potentials. Strong prosodic stress or over-articulation can be exploited for instance to manage emphasis and direct the attention of the listener but also to favour 'acoustic segmentation and lexical access' (Bernabé-Caro & Orero, 2019, p. 66). Using pauses in strategic spots and breathing properly is also conducive to listenability as calibrating the pace of the narration and setting the right tone for each podcast. (p. 458)

When focusing on the analysis of the text, Perego acknowledges the lack of a listenability measure but resorts to a readability formula (Gunning Fog Index) to assess complexity and further analyses the following features: accent, speech rate, pauses, and articulation and emphasis through paralanguage on some words. Perego (2024) also considers that "varying the rate through the speech is essential to make it more interesting and engaging" (p. 458).

4. Recommendations on Easy Spoken Language

Effective communication requires adapting language to meet the diverse needs of audiences, particularly when moving from written texts to spoken formats. While numerous guidelines exist for Easy Language in written texts, recommendations for spoken language remain scarce. This section compiles recommendations addressing unique aspects of oral communication, such as pacing, pronunciation, and pauses, which are less explored compared to their written counterparts. Matamala (2023) provides an overview of such recommendations, which are summarised below:

- Pace: "the speaker should read at a measured pace, neither too fast nor too slow" (Freyhoff et al., 1998), "do not be in a hurry, do not speak too fast" (Inclusion Europe, 2009). The ISO standard 23859 states that a suitable speed should be offered by default, but the user should be able to personalise it.
- Pauses: "[the speaker should] pause between sentences" (Freyhoff et al., 1998) or "at sensible points" (Inclusion Europe, 2009), conveying sentence and text structure (ISO 23859).
- Pronunciation: "make sure that the person speaking had good pronunciation and articulates clearly" (Inclusion Europe, 2009).
- Volume: "not too loud, not too quiet" (Inclusion Europe, 2009). The ISO standard suggests "a default acceptable volume should be provided, giving the user the possibility to adapt it to their needs".
- Accents: "Make sure that the person speaking does not have an accent which is too strong" (Inclusion Europe, 2009).

- Information selection: “for many people retaining too much auditory information is hard” (Accessible Information Working Group, 2011), “[d]o not hesitate to repeat your information several times” (Inclusion Europe, 2009).
- Vocabulary: “Make sure you use words which are easy to understand. For example, do not use dialects” (Inclusion Europe, 2009).
- Emotions: “Read the text in a way that emotions can be perceived” (Inclusion Europe, 2009).
- Voice selection: “When it is appropriate, the voice should match the character” (Inclusion Europe, 2009), “use low-pitched voices, as higher pitcher voices can be difficult to hear for persons with hearing loss” (Accessible Information Working Group, 2011).
- Good quality audio (Accessible Information Working Group, 2011), “without interference or background noise” (Inclusion Europe, 2009), “a good mix” (ISO standard 23859).

As part of the SELSI project (2023), the results of a survey with 446 European respondents were published, assessing what features are important for successful, easy one-way and two-way oral communication. Results show that “when producing simplified content, professionals tend to rely on existing guidelines even if these guidelines are meant for easy written language” (p. 116). One aspect prioritised by respondents is a calm and friendly environment and a respectful and patient attitude, together with eye contact to engage the listeners. The report refers to the “paramount role of pauses for users”, whereas linguistic strategies generally found for written guidelines are mentioned: simple words, explanations, and short sentences. The report adds: “Interestingly, repetition does not seem to be prioritized by professionals despite its positive effect on users” (p. 117).

Finally, Pujadas and Matamala (forthcoming) report on the results of 11 interviews with professionals producing easy oral language content in different formats (easy news, easy interpreting, easy AD) and languages. Results are in line with the SELSI report, as most interviewees find written recommendations appropriate. The specificity is the reference to prosodic features like “slower speech, speed, pauses, and good pronunciation”, aspects that merit further research.

5. Key Features in Listenability

In this section, some key features that enhance listenability will be defined, based on the literature. Some characteristics are shared between spoken and written texts. Adequate grammar and vocabulary are expected to make both spoken and written texts easier. Some suggestions found in the literature (see previous sections) include using clauses of moderate length, coordination rather than subordination, frequent vocabulary, and repetitions. Another shared aspect is good organisation, be it a spoken or a written discourse. For instance, internal summaries and advance organisers are effective strategies for making a text easier to understand. In the field of EL, additional explanations, be them written or spoken, are also recommended. More research is needed on how these strategies affect comprehension across languages and modalities, but there is a consensus that both grammar and vocabulary influence the comprehensibility of spoken and written texts.

The features of the communicative act may also have an impact on the selection of linguistic features. Dialogue differs from one-way information. Recorded audio lacks the adaptability of live speech, where eye contact and engagement strategies can be adjusted. Listening to radio versus audiovisual content with visuals also differs. Key factors include the speakers, context, text genre, speaker attitudes, and acoustic factors like background noise, all of which affect intelligibility and listenability.

Still, regardless of the communicative context, there are some aspects linked to the spoken language to be considered. Next, the focus will be put on aspects related to pronunciation and prosody which are specific to spoken language. Each aspect will be examined to shed light on the concept of listenability.

5.1. Clear Enunciation With a Focus on Pronunciation

Clear enunciation refers to the clarity with which sounds and, obviously, words are pronounced. Poor articulation can affect communication negatively, and the message may not be understood adequately. Lack of awareness and laziness are common challenges associated with articulation. If there is no pathology affecting speech, the complexity in pronunciation is due to the phonotactic constraints of each language. There are two syllabic constituents: onset and rhyme (Easterday, 2019). The onset is the element or elements that precede the nucleus of the syllable (it is usually a vowel or liquid consonant such /r/ or /l/). The rhyme is composed by the nucleus and other glides and consonant elements in the coda. Generally, syllabic structures that have a complex onset or a complex rhyme present greater difficulties in pronouncing (Pimentel et al., 2020). The higher the number of elements in a syllabic unit, the greater the difficulty in articulating them. While elision and errors in pronunciation may be acceptable in spontaneous speech, they can negatively impact speech clarity.

5.2. Appropriate Prosody

Speech prosody involves how someone expresses an utterance to convey emotions, indicate a communicative act, or clarify meaning, and it is crucial for social reciprocity (Bone et al., 2014). Prosody relies on a common set of acoustic cues linked to pitch, voice volume, and voice quality (Roy et al., 2017). Speech rate and pauses also influence prosody. All these aspects are considered next.

5.2.1. Pitch

In relation to pitch, preferences across languages should be considered. For instance, Spanish speakers tend to prefer low-pitched voices which seems to characterise a radiogenic voice in Spanish (Rodero, 2001, 2002), defined as a pleasant voice, harmonious, relaxed, warm and transparent (Rodríguez Bravo, 1989).

Helfrich and Weidenbecher (2011) manipulated three different texts spoken in German with different average pitches. The findings suggest a direct link between pitch and speaker evaluation, revealing that lower-pitched voices received more favourable ratings compared to their higher-pitched counterparts. Guyer et al. (2021) also show the relevance of low voices in the perception of speakers. A low pitch significantly resulted in higher assessments of speaker confidence compared to a high pitch. Furthermore, a low pitch was associated with a greater degree of persuasion than a high pitch. There seems to be a general tendency to select low-pitched voices over high-pitched voices as more trustworthy, both for men and women (Tsantani et al., 2016).

Still in relation to pitch, Machuca and Matamala (2022) aimed to find a perceptual definition of a neutral voice in Catalan, English, and Spanish since some recommendations in audio description indicate that a neutral voice is used. Their study suggests that there is no consensus on the most neutral voice, but participants tend to agree on the less neutral voice. Spanish participants associate non-neutral voices with high-pitch values for both men and women. English participants identify high-pitched female and low-pitched male voices as non-neutral. Catalan participants perceive male low-pitched and female medium-pitched voices as non-neutral. These findings highlight language and gender-specific perceptions of voice neutrality, aspects that could be further explored in relation to listenability.

5.2.2. Volume

The sound pressure level (SPL) refers to the volume of an individual's voice when speaking, i.e. the loudness or softness of our voice. It is a physical measurement in decibels (dB). Volume (also referred to as loudness) is sometimes defined as the subjective perception of the intensity of a sound. There is not much literature on the SPL that is considered standard in a speaker's voice, and even less on SPL related to listenability. The SPL varies depending on the circumstances in which the voice is used. For instance, speaking softly in a quiet environment versus speaking loudly to be heard in a noisy setting. Berg et al. (2017) assessed voice profiles during speech in a group of 2472 individuals (1154 males and 1318 females) aged 40 to 79 from Leipzig (Germany). Their aim was to define standard values for the speaking voice. Four different situations were measured: the softest speaking voice (I), conversational voice (II), classroom voice (III), and shouting voice (IV). According to Berg et al.'s results, values appropriate for listenability are between 52 dB and 68 dB. Voices with volumes below this minimum value might be difficult to understand, and those above the maximum value would come close to shouting, which could be annoying to the listeners.

5.2.3. Voice Quality

Voice quality can be defined as "the characteristic auditory colouring of an individual's voice, derived from a variety of laryngeal and supralaryngeal features and running continuously through the individual's speech" (Trask, 1996, p. 381). A clear and well-modulated voice can help convey the emotions or intentions of the speaker more effectively, leading to better communication. Furthermore, a pleasant voice can capture and maintain the listener's attention, enhancing the overall listening experience and retention of information. For example, Schiller et al. (2023) explored how hoarseness impacts the ability to remember auditory-verbal information, the level of effort required for listening, and the subjective perception of listening quality among adult listeners. Results showed that when someone's voice is hoarse, it might make it harder for people to listen and to understand.

5.2.4. Speech Rate

Speech rate is another important feature in listenability. The ideal speech rate, measured in words per minute, can vary depending on the context, audience, and purpose of the speech and across languages (Tivadar, 2017). Moreover, every listener has their own preferred speech rate, which is observed when they can choose the most appropriate speech rate. Despite this, a common guideline for effective and easily understandable communication is a speech rate of around 150 to 170 wpm. This pace allows listeners to comprehend the spoken content without feeling overwhelmed or struggling to keep up. Different measures are used in speech rate (Tilsen & Tiede, 2023), such as words per minute, syllables per second or sounds per minute or per second, so it is difficult to establish a comparison between these metrics. It seems that the average appropriate pace in syllables per second is between 5.5 and 7 syllables (Tivadar, 2017), which corresponds to ten to twelve phonemes per second (Roach, 2009).

Most research linking speech rate and listening comprehension agrees that comprehension begins to decline at over 250 words per minute (cf. Rubin, 1994). In English, for instance, the range would be between 165 and 180 (Rubin, 1994). In Spanish, Rodero (2016) found that the optimal level of speech rate on the radio was between 170 wpm, with high information density, and 190 wpm, with low information density, with samples under 150 wpm and above 210 wpm being considered difficult to understand. However, a speech rate that is too slow can also hinder understanding by causing listeners to lose interest or become distracted, reducing engagement. Slow speech rate improves listening comprehension in foreign language students (Le 2006, among others), but it is only recommended as a brief exercise to understand natural speech rates (Hayati, 2010).

5.2.5. Pauses

Pauses are key to effective communication, highlighting points and improving listenability by adding structure and processing time. This summary examines silent pauses, focusing on their length, placement, and frequency.

Regarding length, defining a pause's minimum duration is challenging. Goldman-Eisler (1958, 1968) suggest thresholds of 200 ms and 250 ms; different values were found by Llisterri et al. (2024) when they reviewed pause durations. Prafiyanto et al. (2018) found that pauses between 200 and 800 ms did not significantly affect listenability or comprehension, with 200 ms being optimal for faster information transfer.

Location is also key. Prafiyanto (2019) indicates that pauses should align with sentence structure for better comprehension. Estebas-Vilaplana et al. (2023) note that readers pausing at punctuation marks are better perceived in Spanish, highlighting the importance of appropriate pause placement for clarity.

In relation to pause frequency, few pauses can make speech sound rushed, while too many pauses can suggest uncertainty and disengage listeners. Cucchiarini et al. (2002) emphasise that pause frequency, rather than its duration, influences speech smoothness, though an optimal pause frequency is not clearly defined.

6. Conclusions

When researching easy audios, one needs to move from the written text to the spoken word. Written texts need to be legible and readable, focusing on clear formatting and language to enhance comprehension. Similarly, spoken segments need to be intelligible and easy to listen to, considering crucial elements linked to orality. Research and recommendations in language learning, AVT, and accessibility show some shared aspects between written and spoken language, such as sentence structure, vocabulary, and text organisation. However, key aspects unique to oral language also emerge.

Firstly, acoustic factors such as background noise may affect the intelligibility and listenability of a discourse. Secondly, genres and text types in spoken language may affect the strategies to enhance listenability. For example, in a live, spontaneous informal dialogue, one can call the interlocutor by name and use tag questions, while this may not be suitable in a more planned formal setting. Thirdly, prosodic elements are also central. Based on the previous findings, at least the following aspects should be considered:

- articulation,
- pitch,
- volume,

- voice quality,
- speech rate,
- pauses.

One could guarantee a good listening experience with clear articulation, a pleasant pitch in the language being spoken (probably a low-pitch), an adequate voice quality, a standard speech rate within certain ranges, and an adequate number of pauses in the right places. However, it remains to be seen what the expectations would be for easy audios.

Easy audios address users who struggle to understand content, such as persons with intellectual disabilities. In easy audios, slower speech rates and longer pauses could improve listenability, but a critical issue is to avoid listeners losing interest. Further research is needed to establish what works better across languages and users.

Funding Information

This work was supported by the Spanish Ministry of Science, Innovation and Universities (WEL Project, grant number PID2022-137058NB-I00) and by the Secretaria d'Universitats i Recerca del Departament d'Empresa i Coneixement de la Generalitat de Catalunya (grant number 2021SGR00077).

References

Accessible Information Working Group. (2011). *Make it easy: A guide to preparing easy to read information*. Inclusion Ireland. https://inclusionireland.ie/wp-content/uploads/2020/10/makeiteasyguide2011.pdf

Arias-Badia, B., & Matamala, A. (2020). Audio description meets Easy-to-Read and Plain Language: Results from a questionnaire and a focus group in Catalonia. *Zeitschrift für Katalanistik, 33/2020*, 251–270.

Arias-Badia, B., & Matamala, A. (2023). Audio description from an easy-to-understand perspective: A corpus-based study in Catalan. *Jostrans: Journal of Specialised Translation, 40*, 268–296.

Baese-Berk, M. M., Levi, S. V., Van Engen, K. J. (2023). Intelligibility as a measure of speech perception: Current approaches, challenges, and recommendations. *The Journal of the Acoustical Society of America, 153*, 68–76. https://doi.org/10.1121/10.0016806

Beinborn, L., Zesch, T., & Gurevych, I. (2014). Readability for foreign language learning: The importance of cognates. *ITL – International Journal of Applied Linguistics, 165*, 136–162. https://doi.org/10.1075/itl.165.2.02bei

Berg, M., Fuchs, M., Wirkner, K., Loeffler, M., Engel, C., & Berger, T. (2017). The speaking voice in the general population: Normative data and associations to sociodemographic and lifestyle factors. *Journal of Voice, 31*(2), 257. https://doi.org/10.1016/j.jvoice.2016.06.001

Bernabé-Caro, R. (2020). *Easy audiovisual content for all: Easy-to-Read as an enabler of easy, multimode access services* [Unpublished doctoral thesis]. Universitat Autònoma de Barcelona. Research Gate. https://doi.org/10.13140/RG.2.2.22019.94243

Bernabé-Caro, R., & Orero, P. (2019). Easy to read as multimode accessibility service. *Hermēneus, 21*, 53–74.

Bernabé-Caro, R., & Orero, P. (2020). Easier audio description: A more readily understood accessibility service. In S. Braun & K. Starr (Eds.), *Innovation in audio description research* (pp. 55–75). Routledge.

Bone, D., Lee, C. C., Black, M. P., Williams, M. E., Lee, S., Levitt, P., & Narayanan, S. (2014). The psychologist as an interlocutor in autism spectrum disorder assessment: Insights from a study of spontaneous prosody. *Journal of Speech, Language, and Hearing Research, 57*(4), 1162–1177. https://doi.org/10.1044/2014_JSLHR-S-13-0062

Chiu, C., & Chen, T. (2023). Speech rate and young EFL learners' listening comprehension. *English Language Teaching, 16*(7), 74–80.

Cruz, N. C. (2007). Terminologies and definitions in the use of intelligibility: State-of-the-art. *Revista Brasileira de Linguística Aplicada, 7*(1), 149–159.

Cucchiarini, C., Strik, H., & Boves, L. (2002). Quantitative assessment of second language learners' fluency: Comparisons between read and spontaneous speech. *The Journal of the Acoustical Society of America, 111*, 2862–2873. https://doi.org/10.1121/1.1471894

Denbow, C. J. (1975). Listenability and readability: An experimental investigation. *Journalism & Mass Communication Quarterly, 52*(2), 285–290. https://doi.org/10.1177/107769907505200213

DuBay, W. H. (2007). *Smart language: Readers, readability, and the grading of text*. Impact Information.

Easterday, S. (2019). *Highly complex syllable structure: A typological and diachronic study (Studies in Laboratory Phonology 9).* Language Science Press.

Eastwood, J., & Snook, B. (2012). The effect of listenability factors on the comprehension of police cautions. *Law and Human Behavior, 36*(3), 177–183. https://doi.org/10.1037/h0093955

Estebas-Vilaplana, E., Garrido, J. M., & Machuca, M. J. (2023). Pausas en la lectura de textos expresivos: Métricas de evaluación [Pauses in expressive text reading: Assessment metrics]. *Langue(s) & Parole, 8*, 59–78. https://doi.org/10.5565/rev/languesparole.129

Flesch, R. (1948). A new readability yardstick. *Journal of Applied Psychology, 32*(3), 221–233. https://doi.org/10.1037/h0057532

Friedrich, M., & Heise, E. (2022). The influence of comprehensibility on interest and comprehension. *Zeitschrift für Pädagogische Psychologie*, 1–14. https://doi.org/10.1024/1010-0652/a000349

Glenn, E. C., Emmert, P., & Emmert, V. (1995). A scale for measuring listenability: The factors that determine listening ease and difficulty. *International Journal of Listening, 9*(1), 44–61.

Goldman-Eisler, F. (1958). Speech production and the predictability of words in context. *Quarterly Journal of Experimental Psychology, 10*(2), 96–106. https://doi.org/10.1080/17470215808416261

Goldman-Eisler, F. (1968). *Psycholinguistics: Experiments in spontaneous speech*. Academic Press.

Gunning, R. (1952). *The technique of clear writing*. McGraw-Hill.

Guyer, J., Briñol, P., & Vaughan-Johnston, T. (2021). Paralinguistic features communicated through voice can affect appraisals of confidence and evaluative judgments. *Journal of Nonverbal Behavior, 45*, 479–504. https://doi.org/10.1007/s10919-021-00374-2

Handayani, G., & Wirza, Y. (2020). An analysis on language content and readability level of primary English textbook. *Advances in Social Science, Education and Humanities Research, 546*, 203–208.

Harwood, K. A. (1955). Listenability and rate of presentation. *Speech Monographs, 22*, 57–59.

Hayati, A. (2010). The effect of speech rate on listening comprehension of EFL learners. *Creative Education, 2*, 107–114. https://doi.org/10.4236/ce.2010.12016

Helfrich, H., & Weidenbecher, P. (2011). Impact of voice pitch on text memory. *Swiss Journal of Psychology, 70*(2), 85–93.

Horowitz, R., & Samuels, S. J. (1985). Reading and listening expository text. *Journal of Reading Behavior, 17*(3), 185–198.

Inclusion Europe. (2009). *Information for all: European standards for making information easy to read and understand*. Inclusion Europe. https://inclusion-europe.eu/wp-content/uploads/2015/03/2113-Information-for-all-16.pdf

International Organization for Standardization. (2023). ISO/IEC 23859:2023 Information technology – User interfaces – Requirements and recommendations on making written text easy to read and understand. ISO.

Freyhoff, G., Hess, G., Kerr, L., Menzel, E., Tronbacke, B., & Van Der Veken, K. (1998). *Make it Simple. European guidelines for the production of easy-to-read information for people with learning disability*. International League of Societies for the Mentally Handicapped (ILSMH). https://repositori.lecturafacil.net/sites/default/files/1998%20El%20cam%C3%AD%20m%C3%A9s%20f%C3%A1cil%20Directrius%20europees%20per%20gernar%20informaci%C3%B3%20de%20Lectura%20F%C3%A0cil%20ILSMH_ANG.pdf

Karas, H. (2020). Intelligibility and the reception of translations. *Perspectives, 28*(1), 24–42. https://doi.org/10.1080/0907676x.2019.1612929

Kotani, K., Ueda, S., Yoshimi, T., & Nanjo, H. (2014). A listenability measuring method for an adaptive computer-assisted language learning and teaching system. In *Proceedings of 28th Pacific Asia Conference on Language, Information and Computation* (pp. 387–394).

Le, F. (2006). *Faster, normal or slower? The effects of speech rates on high-intermediate ESL learners' listening comprehension of academic lectures* [Unpublished doctoral dissertation]. Iowa State University.

Lindholm, C., & Vanhatalo, U. (2021). *Handbook of easy language in Europe*. Frank & Timme.

Llisterri, J., Machuca, M. J., & Ríos, A. (2024). Algunas reflexiones sobre el estudio de las pausas silenciosas. In W. Elvira-García & P. Roseano (Coords.), *Avances metodológicos en fonética y prosodia* [Methodological advances in phonetics and prosody] (pp. 145–155). Servicio de Publicaciones de la UNED.

Loukina, A., Yoon, S., Sakano, J., Wei, Y., & Sheehan, K. (2016). Textual complexity as a predictor of difficulty of listening items in language proficiency tests. In *Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical papers* (pp. 3245–3253).

Machuca, M. J., & Matamala, A. (2022). Neutral voices in audio descriptions: What does it mean? *Babel, 68*(5), 668–696. https://doi.org/10.1075/babel.00287.mac

Matamala, A. (2019). *Accessibilitat i traducció audiovisual*. Vol. 24, Biblioteca de Traducció i Interpretació. Eumo Editorial. https://doi.org/10.24197/her.23.2021.539-543

Matamala, A. (2022). Easy-to-understand language in audiovisual translation and accessibility: State of the art and future challenges. *X-Linguae, 15*(2), 130–144. https://doi.org/10.18355/XL.2022.15.02.10

Matamala, A. (2023). Easy audios for easy audiovisual content: An overview. In C. Pena Díaz (Ed.), *The making of accessible audiovisual translation* (pp. 83–106). Peter Lang.

Messerklinger, J. (2006). Listenability. *Center for English Language Education Center, 14*, 56–70.

Nahón Guillén, M. (2020). Simultaneous simplification: A world first, to include persons with intellectual disabilities. *Communicate Journeys, 76*. https://aiic.org/site/webzine/issue-76/simultaneous-simplification

Navetta, A. (2021). Easy-to-understand (E2U) and audio description (AD). Processes interview with professionals: Listenability. *EASIT Project*. https://ddd.uab.cat/pub/recdoc/2021/245293/U3B.E1._Interview_with_professionals_-_Listenability_UNITS_TRANSCRIPT.pdf

Oncins, E., Bernabé-Caro, R., Montagut, M., & Arnáiz-Uzquiza, V. (2020). Accessible scenic arts and virtual reality: A pilot study in user preferences when reading subtitles in immersive environments. *MonTi, 12*, 214–241.

Orero, P. (2022). Audio description personalization. In C. Taylor & E. Perego (Eds.), *The Routledge handbook of audio description* (pp. 121–134). Routledge.

Perego, E. (2023). *Audio description for the arts: A linguistic perspective*. Routledge.

Perego, E. (2024). Translation into easy language: The unexplored case of podcasts. In L. Pillière & Ö. Berk Albachten (Eds.), *The Routledge handbook of intralingual translation* (pp. 453–471). Routledge.

Perego, E., & Taylor, C. (Eds.). (2022). *The Routledge handbook of audio description*. Routledge.

Pimentel, T., Roark, B., & Cotterell, R. (2020). Phonotactic Complexity and Its Trade-offs. *Transactions of the Association for Computational Linguistics*, 8 1–18. https://doi.org/10.1162/tacl_a_00296

Prafiyanto, H. (2019). *Study on presentation and evaluation of speech toward development of a Japanese oral communication training system* [Unpublished doctoral dissertation]. Tohoku University.

Prafiyanto, H., Nose, T., Chiba, Y., & Ito, A. (2018). Analysis of preferred speaking rate and pause in spoken easy Japanese for non-native listeners. *Acoustical Science and Technology, 39*(2), 92–100. https://doi.org/10.1250/AST.39.92

Pujadas, M., & Matamala, A. (forthcoming). Creation and translation of audiovisual content in oral easy-to-understand language: Insights from interviews with professionals. *Hermēneus*, 27.

Roach, P. (2009). *English phonetics and phonology: A practical course*. Cambridge University Press.

Rodero, E. (2001). *Locución informativa radiofónica* [Broadcast news delivery] [Doctoral dissertation]. Universidad Pontificia de Salamanca.

Rodero, E. (2002). El tono de la voz masculina y femenina en los informativos radiofónicos: Un análisis comparativo [The tone of male and female voices in radio news: A comparative analysis]. In *Proceedings of Mujeres, Hombres y Medios de Comunicación* (pp. 319–331). Lex Nova.

Rodero, E. (2016). *Elementos prosódicos para mejorar la memoria en la publicidad radiofónica: Análisis del efecto de diferentes estrategias de entonación, acento y velocidad de habla sobre el recuerdo de mensajes publicitarios en radio* [Prosodic elements to improve memory in radio advertisement: An analysis of the effect of different intonation, accent and speed strategies on the recall of radio adverts] [Unpublished doctoral dissertation]. Universitat Autònoma de Barcelona. https://ddd.uab.cat/record/166005

Rodríguez Bravo, A. (1989). *La construcción de una voz radiogénica* [The construction of a broadcast-ready voice] [Unpublished doctoral dissertation]. Universitat Autònoma de Barcelona.

Roy, J., Cole, J., & Mahrt, T. (2017). Individual differences and patterns of convergence in prosody perception. *Laboratory Phonology, 8*(1), 22. https://doi.org/10.5334/labphon.108

Rubin, D. L. (1993). Listenability = Oral-based discourse + Considerateness. In A. D. Wolvin & C. G. Coakley (Eds.), *Perspectives on listening* (pp. 261–281). Ablex Publishing.

Rubin, J. (1994). A review of second language listening comprehension research. *The Modern Language Journal, 78*(2), 199–221.

Rubin, D. L. (2012). Listenability as a tool for advancing health literacy. *Journal of Health Communication, 17*(sup3), 176–190. https://doi.org/10.1080/10810730.2012.712622

Schiller, I. S., Aspöck, L., & Schlittmeier, S. J. (2023). The impact of a speaker's voice quality on auditory perception and cognition: A behavioral and subjective approach. *Frontiers in Psychology, 14*, 1243–1249. https://doi.org/10.3389/fpsyg.2023.1243249

SELSI Project. (2023). *Spoken easy language for social inclusion*. https://selsi.eu/wp-content/uploads/2023/10/WP2-Report_FINAL.pdf

Sticht, T., & James, J. (1984). Listening and reading. In P. Pearson (Ed.), *Handbook of research on reading* (Vol. 2, pp. 293–317). Longmans.

Tilsen, S., & Tiede, M. (2023). Parameters of unit-based measures of speech rate. *Speech Communication, 150*, 73–97. https://doi.org/10.1016/j.specom.2023.05.006

Tivadar, H. (2017). Speech rate in phonetic-phonological analysis of public speech (using the example of political and media speech). *Jazykovedný časopis, 68*(1), 37–56. https://doi.org/10.1515/jazcas-2017-0016

Trask, R. L. (1996). *A dictionary of phonetics and phonology*. Routledge.

Tsantani, M. S., Belin, P., Paterson, H., & McAleer, P. (2016). Low vocal pitch preference drives first impressions irrespective of context in male voices but not in female voices. *Perception, 45*(8), 946–963.

Weaver, C. H. (1972). *Human listening: Processes and behavior*. The Bobbs-Merrill Company.

WEL project. (2023–2026). *WEL (From written to oral texts in easy language: Easy audios in cultural visits and video games)*. https://webs.uab.cat/wel/en/

Wolfer, S. (2015). Comprehension and comprehensibility. In K. Maksymski, S. Gutermuth, & S. Hansen-Schirra (Eds.), *Translation and comprehensibility*. Frank & Timme.

Xia, M., Kochmar, E., & Briscoe, T. (2016). Text readability assessment for second language learners. In *Proceedings of the 11th Workshop on Innovative Use of NLP for Building Educational Applications* (pp. 12–22).

Yalon-Chamovitz, S., & Avidan-Ziv, O. (2016). Simultaneous simplification: Stretching the boundaries of UDL. *Preconference Universal Design for Learning Implementation and Research Network 2016*. https://adaptit.co.il/wp-content/uploads/2020/01/Simultaneous-Simplification-Stretching-the-Boundaries-of-UDL.pdf

Zhao, Y. (1997). The effects of listeners' control of speech rate on second language comprehension. *Applied Linguistics, 18*(1), 49–68.