UDC 37.091.3:004.032.6 **DOI https://doi.org/10.24195/2617-6688-2026-1-3**

Zembytska Maryna
PhD in Pedagogy (Candidate of Pedagogical Sciences), Associate Professor
Khmelnytskyi National University, Khmelnytskyi, Ukraine
E-mail: marynazembytska@gmail.com
ORCID ID: https://orcid.org/0000-0002-6671-9937
Researcher ID: HIK-0164-2022
Scopus: 57865200700

Using Video-Based Corpora to Enhance EFL Students' Connected Speech and Pronunciation

The relevance of this research arises from the well-known difficulty EFL learners face in mastering connected speech and suprasegmental features, such as assimilation, elision, linking, vowel reduction, rhythm, stress, and intonation, which are essential for achieving fluent, intelligible spoken English in real-life communication. Traditional phonetics instruction in university settings often prioritizes isolated segmental practice (individual sounds via minimal pairs and drills), while largely neglecting suprasegmentals and the natural reductions, liaisons, and prosodic patterns characteristic of authentic rapid speech. This mismatch leaves many intermediate-to-advanced learners (B2 – C1 levels) struggling with listening comprehension and production in unscripted contexts, despite solid knowledge of citation forms. In globalized EFL environments, where English functions as a lingua franca with diverse accents and high speech rates, prioritizing intelligibility over native-like accuracy has become a widely accepted pedagogical goal. Video-based corpora – collections of authentic audiovisual speech from platforms like YouGlish, TED Talks, and BBC Learning English – offer rich, multimodal, contextualized input that can bridge this gap by exposing learners to genuine phonetic variation, visual articulatory cues, and discourse-level meaning, thereby supporting more effective, learner-centered pronunciation development.

The purpose of the article is to investigate and demonstrate the pedagogical effectiveness of integrating video-based corpora into structured EFL pronunciation instruction, specifically to enhance learners' perceptual awareness, production accuracy, and confidence in handling connected speech phenomena and suprasegmental features. The study evaluates a guided three-stage instructional framework (perception – analysis – production) combined with data-driven learning (DDL) principles, assessing its impact on a group of university students and deriving practical recommendations for phonetics course design.

Research methods combined a small-scale, classroom-based mixed-methods approach. The participants were 24 voluntary B2 – C1 undergraduate EFL students at Khmelnytskyi National University enrolled in a Practical Phonetics course with minimal prior connected speech training. Over six weeks (12 sessions), learners engaged with authentic video materials from YouGlish (for targeted phrase/accent searches), TED Talks (extended discourse with multimodal cues), BBC Learning English (graded examples with transcripts), and other sources. The intervention followed a structured sequence: perception tasks (identifying features in scaffolded clips), analysis tasks (transcription, annotation, inductive pattern discovery via DDL), and production tasks (shadowing, repetition, self-recording, and communicative role-plays). Pre- and post-intervention assessments included perception tests (identification of linking, intrusive sounds, weak forms) and production tasks (recorded sentences rated for accuracy, prosodic consistency, and fluency). Quantitative data were analyzed using descriptive statistics and significance testing; qualitative insights came from learner reflections, open-ended questionnaires, classroom observations, and thematic analysis, with triangulation ensuring validity. Ethical protocols (informed consent, anonymity) were observed.

The research details theoretical foundations (Krashen's comprehensible input hypothesis, multimodal learning theory, DDL, intelligibility-oriented pronunciation pedagogy), a rationale for shifting from decontextualized drills to authentic audiovisual corpora, and a practical three-stage framework. Illustrative DDL activities are presented: querying YouGlish for function-word reductions ("going to" → /gənə/), linking/intrusive /r/ patterns, assimilation in clusters ("handbag" → /hæmbæg/), and elision in consonant sequences ("next please" → /neks pliːz/). These tasks foster inductive discovery, metalinguistic awareness, and autonomy. Empirical evidence includes pre/post comparative tables showing statistically significant gains across targeted features.

The results indicate substantial improvements: perceptual identification of linking rose from 52 to 86% (+34%, $p < 0,05$), intrusive sounds from 41 to 78% (+37%, $p < 0,01$), overall production accuracy from 48 to 72% (+24%, $p < 0,05$), prosodic consistency from 45 to 70% (+25%, $p < 0,05$), nuclear stress from 58 to

89% (+31%), and vowel reduction in weak forms from 39 to 81% (+42%). Qualitative data revealed increased phonetic awareness, reduced speaking/listening anxiety, greater confidence in authentic contexts, and positive perceptions of video input ("seeing the speaker helped me understand reductions"). The findings confirm that guided, multimodal exposure to video corpora, scaffolded through perception-analysis-production cycles and DDL, significantly outperforms traditional deductive methods for suprasegmental and connected speech instruction. The article discusses implementation challenges (cognitive overload from speed/accent variation, need for curation/scaffolding, technical access), argues for broader adoption in university phonetics curricula to promote intelligibility and learner autonomy, and outlines directions for future research, including longitudinal studies, larger/diverse samples, integration with automatic speech recognition, and comparisons across proficiency levels or corpus types.

Keywords: *EFL teaching, pronunciation, connected speech, suprasegmental features, phonetic competence, video-based corpora, comprehensible input, data-driven learning, multimodal learning.*

Introduction. Pronunciation is known to be a cornerstone of communicative competence in a foreign language. In EFL contexts, effective spoken interaction requires not only accurate articulation of individual sounds but also mastery of natural connected speech patterns, such as assimilation (e.g., "handbag" as /hæmbæg/), elision (e.g., "next please" as /neks pliːz/), linking (e.g., "an apple" as /əˈnæpl/), reduction (e.g., "can" as /k(ə)n/), sentence stress, rhythm, and intonation. These features are essential for fluent communication in real-world settings, where speech is rapid and variable.

Suprasegmental and connected speech features are frequently underrepresented in phonetic courses, which often prioritize segmental accuracy (e.g., individual vowels/consonants) via isolated drills and minimal-pair exercises. As a result, learners may perform well in the classroom but struggle with listening comprehension and fluency in authentic discourse.

In this context, video-based corpora, featuring naturally occurring audiovisual speech from real speakers, offer a powerful tool for delivering authentic, comprehensible input. Using such resources, teachers enable learners to observe how phonetic features interact with communicative meaning, bridging the gap between theoretical knowledge and practical use of language.

Previous research has consistently demonstrated that successful oral communication in a foreign language depends not only on segmental accuracy but, to a greater extent, on control of suprasegmental features and connected speech processes. Early works by (Brazil, 1997) and (Celce-Murcia et al., 2010) established stress, rhythm, and intonation as core components of intelligible speech, while later empirical studies confirmed that weak forms, vowel reduction, assimilation, elision, and linking play a crucial role in speech fluency and listener's comprehension (Derwing & Munro, 2015; Levis, 2018). In contemporary EFL teaching, the focus has shifted from native-like norms toward intelligibility and comprehensibility as pedagogically sound goals, particularly in EFL contexts (Jenkins, 2000; Jenkins et al., 2018; Galante & Piccardo, 2022). Research indicates that learners who receive explicit instruction in suprasegmentals demonstrate improved listening accuracy, more natural rhythm, and greater communicative confidence, even when segmental deviations remain (Hahn, 2004; Gordon & Darcy, 2016).

In recent decades, corpus-based approaches have been proposed as a means of exposing learners to authentic spoken language patterns. Data-driven learning (DDL), first described by Johns (1991), positions students as active analysts who learn about linguistic regularities through systematic observation of corpus data. While DDL has been extensively applied to grammar and vocabulary, its application to pronunciation and connected speech has come into the spotlight only recently (Leńko-Szymańska, 2017; Ma, 2024). Studies suggest that inductive corpus exploration is more likely to enhance phonological awareness than rote memorization (Gilquin & Granger, 2010; Boulton & Cobb, 2017). When corpora include spoken or audiovisual data, learners are better able to perceive reductions, stress shifts, and rhythmic variation that are typically lost in written representations of speech (Cauldwell, 2013).

Video-based corpora extend pronunciation learning by incorporating visual information that supports auditory perception. Research on multimodal learning theory demonstrates that the integration of visual and auditory information enhances phonological processing, particularly for abstract prosodic features (Pujolà, 2002; Mayer, 2009).

In pronunciation instruction, visual cues such as lip movement, jaw relaxation, head nodding, and gesture have been shown to facilitate perception of stress, intonation, and vowel reduction (Hardison, 2004; Sueyoshi & Hardison, 2005). Empirical studies report positive effects of video-based pronunciation instruction on learners' fluency, prosodic control, and listening comprehension, particularly when tasks are scaffolded and analytically guided (Saito & Plonsky, 2019; Alisoy, 2025).

Formulation of the problem. Despite their pedagogical potential, the literature suggests that unguided exposure to authentic video input may be overwhelming to learners due to rapid speech rates and accent variability (Field, 2008; Vanderplank, 2016). Consequently, there's a need for structured instructional models that combine perception, analysis, and production phases, allowing learners to notice, test, and adopt phonetic patterns gradually (Celce-Murcia et al., 2010; Gordon, Darcy, & Ewert, 2013). The integration of video-based corpora aligns with Krashen's (1985) notion of comprehensible input and supports learner autonomy while maintaining pedagogical control. Nevertheless, empirical research examining the systematic use of video-based corpora specifically for connected speech instruction at the university level remains limited. The present study seeks to contribute to this field by investigating how guided use of video corpora can enhance EFL learners' perception and production of connected speech features.

This small-scale mixed-methods classroom study involved 24 voluntary B2 – C1 undergraduates at Khmelnytskyi National University in a Practical Phonetics course. Participants had limited prior instruction in connected speech. Over six weeks (12 lessons), video-based corpora were integrated with perception, analysis, and production tasks using YouGlish, TED Talks, interviews, and lectures. The participants gave informed consent, and their anonymity was ensured. Pre- and post-tests measured perception (identification tasks) and production (recorded sentences rated for accuracy and fluency). Qualitative data included reflections, open-ended questionnaires, and observations. Descriptive statistics analyzed quantitative results; thematic analysis handled qualitative data, with triangulation (multiple sources) for validity.

The main part. This study builds on second language acquisition (SLA) principles, multimodal learning theory, and contemporary pronunciation teaching. Krashen's (1985) comprehensible input hypothesis claims that acquisition occurs most effectively through exposure to input slightly beyond current competence (i+1) yet understandable via context. In pronunciation teaching, this means providing authentic spoken language that is largely comprehensible while embedding target phonetic features. Video input supports this by offering rich contextual cues (e.g., situational visuals, gestures), allowing learners to focus on form without semantic overload.

Multimodal learning theory emphasizes that combining auditory cues with visual elements (gestures, facial expressions, articulatory movements) facilitates deeper processing. For EFL learners, audiovisual input enhances perception of suprasegmentals – features like stress timing or vowel reduction that are abstract and hard to grasp from audio alone. Visuals provide cues (e.g., jaw relaxation during schwa production), which make it easier for learners to notice and adopt certain pronunciation patterns.

Video-based corpora (audiovisual collections of real speech, often with transcripts) differ from scripted textbook materials by offering natural variation in different speech rates, accents, and discourse contexts. Platforms, such as YouGlish (for targeted lexical searches), TED Talks (meaningful discourse), and BBC Learning English (contextualized examples) enable repeated, focused exposure to suprasegmental features, speaker differences, and accent varieties.

A DDL approach empowers learners as "language detectives", using corpora to discover patterns rather than memorizing rules. For instance, learners search a corpus for occurrences of "to" in connected speech, observe frequent reduction to /tə/, and hypothesize vowel reduction based on empirical evidence across contexts. This fosters high-level pattern recognition, metalinguistic awareness, and autonomy, shifting authority from instructor to data. Linking /r/ (e.g., "four apples") can be demostrated similarly: learners query YouGlish, analyze 10+ realizations across speakers, and observe articulatory reactivation in video (tongue movement bridging vowels).

In this study, we used a three-stage framework to incorporate video-based corpora in the course of phonetics:

1. Perception: learners viewed short excerpts (e.g., 30–60 seconds) to identify connected speech features (weak forms, linking, rhythm etc.). Tasks included marking reductions on transcripts or noting stress patterns while watching. For lower-proficiency segments, slower BBC clips provided initial entry points. Repeated viewing was facilitated so that segmentation skills were gradually developed and dependence on citation forms was reduced.

2. Analysis: learners transcribed excerpts, annotated assimilation/elision/stress, and compared with captions or visualizations. This was supposed to link theory to authentic examples, encouraging inductive discovery (e.g., searching YouGlish for weak and strong forms of *"and"*).

3. Production: This stage included both reproductive and productive pronunciation practice. Learners first reproduced the phonetic patterns observed in the video corpus through shadowing and controlled repetition, mimicking the timing, prosody, stress patterns, and reductions occurring in the original recordings. These activities helped students internalize connected speech features and align their articulation with authentic models. Subsequently, learners moved to guided communicative production tasks, such as self-recorded sentence production, short role-plays, and mini-dialogues, in which they incorporated weak forms, linking, assimilation and appropriate stress patterns. Students compared their recordings with the corpus models and reflected on differences, combining model-based reproduction with independent communicative production.

This instructional sequence balances various platforms to enhance specific skills: YouGlish facilitates exposure to diverse regional accents, while TED Talks provide rich context for prosodic analysis through multimodal cues like gestures. BBC materials are used for initial perception, with transcripts serving as models for phonetic reductions. By incorporating DDL elements, such as independent corpus queries for homework, the sequence fosters learner autonomy.

The intervention resulted in statistically significant improvements across all targeted phonetic features. As shown in Table 1, the most dramatic increase in perception was noted in Intrusive Sound Perception, which rose by 37%.

For sentence stress, learners contrasted examples (e.g., stress shifts in “I didn’t say he stole the money”) to detect duration, pitch, and reduction. The impact on suprasegmentals was even more pronounced (Table 2). Vowel Reduction in Weak Forms saw a 42% improvement, suggesting that the DDL approach helped students perceive reductions as rhythmic necessities.

Within this framework, data-driven learning principles are applied so that learners assume the role of language researchers. Rather than being presented with prescriptive rules, learners are encouraged to explore video-based corpora inductively. Authentic speech samples are queried and patterns are observed across multiple speakers and contexts. Several illustrative applications of this approach are described below.

When the reduction of function words is investigated, learners are directed to search platforms such as YouGlish for high-frequency phrases including “going to,” “want to”, or “out of”. Multiple video clips are examined, the target words are transcribed, and vowel realizations are noted. Frequently, the vowel is observed to reduce to a schwa in unstressed positions. By collecting and categorizing numerous instances, learners are able to hypothesize that function words tend to undergo vowel reduction in connected speech in order to maintain the characteristic stress-timed rhythm of English. These observations are subsequently tested through shadowing of selected clips, followed by self-recording and comparison with the original models.

In the exploration of linking and intrusive sounds, vowel-final words followed by vowel-initial words are searched (for example, “go away”, “see it”, “idea of”). Instances of linking consonants or intrusive /r/, /w/, or /j/ are identified across various accents. Transcripts are annotated, and patterns are grouped according to accent type (rhotic versus non-rhotic). Learners are thus led to conclude that intrusive /r/ is frequently inserted in non-rhotic varieties, facilitating smooth articulation and preserving rhythmic flow.

Assimilation processes can be examined in a similar manner. Common consonant clusters (such as those found in “handbag”, “ten pens”, or “this shop”) are searched, transcribed, and analyzed. Changes in place, manner, or voicing are documented across different speech rates. Through systematic comparison, learners are able to deduce the conditions under which assimilation occurs – most notably, alveolar stops assimilating to bilabial position before bilabial consonants. Role-play activities incorporating the observed assimilated forms are subsequently conducted, with recordings being produced and assessed for naturalness.

Elision phenomena can also be investigated inductively. Phrases containing potential consonant cluster reduction (for example, “next please”, “last night”, “comfortable”) are retrieved from the corpora. Omitted sounds are marked during transcription, and patterns are identified across contexts. Learners are guided to recognize that medial consonants in three-consonant clusters are frequently elided to maintain articulatory ease. These discoveries are applied in shadowing exercises and mini-presentations, during which learners produce the phrases and monitor their own performance against authentic models.

Table 1

Comparative Analysis of Student Phonetic Performance

Phonetic Feature	Pre-Intervention (Mean %)	Post-Intervention (Mean %)	Improvement Gap	Significance (p-value)
Linking Identification	52%	86%	+34%	<0,05
Intrusive Sound Perception	41%	78%	+37%	<0,01
Production Accuracy	48%	72%	+24%	<0,05
Prosodic Consistency	45%	70%	+25%	<0,05

Table 2

Student Proficiency in Sentence Stress and Prosodic Focus

The Aspect of Connected Speech:	Pre-Intervention	Post-Intervention	Improvement Gap
Nuclear Stress Identification	58%	89%	+31%
Content vs. Function Words Distinction	62%	84%	+22%
Rhythmic Isochrony (Production)	43%	74%	+31%
Vowel Reduction in Weak Forms	39%	81%	+42%

Through these data-driven activities, learners are not merely exposed to isolated rules but are enabled to construct their own understanding of connected speech mechanisms. Metalinguistic awareness is thereby fostered, learner autonomy is promoted, and the internalization of suprasegmental and connected speech features is facilitated more effectively than through traditional deductive instruction.

Discussion. The results confirm that guided, multimodal exposure to video corpora significantly outperforms traditional deductive methods. Qualitatively, students reported reduced anxiety and heightened awareness, noting that "seeing the speaker" provided essential articulatory cues (e.g., jaw relaxation) that audio-only input lacks. These findings align with the "Intelligibility Principle", suggesting that corpus discovery helps students move from L1-influenced syllable-timed speech toward the stress-timed rhythm of English. While challenges such as cognitive overload and accent variability persist, they can be effectively mitigated through the scaffolded framework proposed in this study.

Conclusions. Integrating video-based corpora into university EFL instruction provides a bridge between theoretical phonetics and real-world communication. By combining DDL with a perception-analysis-production cycle, educators can foster greater learner autono my and phonetic competence. However, challenges persist: varied speech rates/overlaps may overwhelm beginners; accent diversity can confuse without core consolidation; inconsistent feature occurrence requires careful curation; multimodal processing risks cognitive overload; technical issues (e.g., captions) and potential idiosyncratic patterns demand scaffolding. Independent home use offers extended exposure but needs guided tasks (transcription, self-recording) to maximize benefits and minimize mislearning.

Future research should pursue longitudinal studies and explore the integration of Automatic Speech Recognition (ASR) tools to further enhance the self-assessment phase of this framework.

References

Akram, M., & Qureshi, A. H. (2014). The role of features of connected speech in teaching English pronunciation. *International Journal of English and Education*, *3*(3), 230–240. Retrieved from https://ijee.org/assets/docs/23.184151609.pdf

Alisoy, H. (2025). The role of using authentic videos on learners' pronunciation. *Acta Globalis Humanitatis et Linguarum*, *2*(2), 49–57. https://doi.org/10.69760/aghel.025002088

Brazil, D. (1997). *The communicative value of intonation in English*. Cambridge University Press.

Cauldwell, R. (2013). *Phonology for listening*. Speech in Action.

Celce-Murcia, M., Brinton, D. M., Goodwin, J. M., & Griner, B. (2010). *Teaching pronunciation: A course book and reference guide* (2nd ed.). Cambridge University Press.

Derwing, T. M., & Munro, M. J. (2015). *Pronunciation fundamentals: Evidence-based perspectives for L2 teaching and research*. John Benjamins. https://doi.org/10.1075/lllt.42

Field, J. (2008). *Listening in the language classroom*. Cambridge University Press.

Galante, A., & Piccardo, E. (2022). Teaching pronunciation: Toward intelligibility and comprehensibility. *ELT Journal*, *76*(3), 375–386.

Gordon, J., & Darcy, I. (2016). The development of comprehensible speech. *Applied Linguistics*, *37*(5), 593–617. https://doi.org/10.1075/jslp.2.1.03gor

Hahn, L. D. (2004). Primary stress and intelligibility: Research to motivate the teaching of suprasegmentals. *TESOL Quarterly*, *38*(2), 201–223. https://doi.org/10.2307/3588378

Hardison, D. M. (2004). Generalization of computer-assisted prosody training: Quantitative and qualitative findings. *Language Learning & Technology*, *8*(1), 34–52. Retrieved from https://scholarspace.manoa.hawaii.edu/server/api/core/bitstreams/274d274e-5483-4ee2-b242-d828985a0fce/content

Jenkins, J. (2000). *The phonology of English as an international language*. Oxford University Press.

Jenkins, J., Baker, W., & Dewey, M. (Eds.). (2018). *The Routledge handbook of English as a lingua franca*. Routledge.

Johns, T. (1991). From printout to handout: Grammar and vocabulary teaching in the context of data-driven learning. *English Language Research Journal*, *4*, 27–45.

Krashen, S. (1985). *The input hypothesis: Issues and implications*. Longman.

Leńko-Szymańska, A. (2017). Training teachers in data-driven learning: Tackling the challenge. *Language Learning & Technology*, *21*(3), 217–241. https://doi.org/10.64152/10125/44628

Levis, J. M. (2018). *Intelligibility, oral communication, and the teaching of pronunciation*. Cambridge University Press. Retrieved from https://dokumen.pub/intelligibility-oral-communication-and-the-teaching-of-pronunciation-1108416624-9781108416627.html

Ma, Q. (2024). Exploring EFL students' pronunciation learning supported by corpus-based tools. *Journal of Language Teaching and Research*, *15*(4), 1029–1038. https://doi.org/10.1080/09588221.2024.2432965

Mayer, R. E. (2009). *Multimedia learning* (2nd ed.). Cambridge University Press. https://doi.org/10.1017/CBO9780511811678

Saito, K., & Plonsky, L. (2019). Effects of second language pronunciation instruction: A meta-analysis. *Studies in Second Language Acquisition*, *41*(3), 497–529. https://doi.org/10.1111/lang.12345

Sueyoshi, A., & Hardison, D. M. (2005). The role of gestures and facial cues in second language listening comprehension. *Language Learning*, *55*(4), 661–699. https://doi.org/10.1111/j.0023-8333.2005.00320.x

Vanderplank, R. (2016). *Captioned media in foreign language learning and teaching*. Palgrave Macmillan. https://doi.org/10.1057/978-1-137-50045-8

Використання відеокорпусів для покращення зв'язного мовлення та вимови студентів, які вивчають англійську як іноземну мову

Зембицька Марина Володимирівна
кандидат педагогічних наук, доцент
Хмельницького національного університету,
Хмельницький, Україна

Актуальність дослідження зумовлена об'єктивними труднощами, з якими стикаються студенти під час опанування спонтанного англійського мовленням. Традиційне навчання в закладах вищої освіти часто фокусується на відпрацюванні окремих звуків, ігнорує водночас надсегментні одиниці (асиміляцію, елізію, ритм, наголос тощо), що призводить до низького рівня розуміння автентичного мовлення. Метою статті є обґрунтування та практична перевірка методики використання відеокорпусів як інструменту вдосконалення фонетичної компетентності здобувачів освіти.

У роботі використано комплекс методів: кабінетне дослідження теоретичних засад (гіпотеза «зрозумілого входу», мультимодальне навчання), педагогічний експеримент із залученням 24 студентів рівня B2 – C1, а також статистичний аналіз результатів. Основним інструментом стали відеоплатформи YouGlish, TED Talks та BBC Learning English.

Зміст дослідження розкриває триетапну модель навчання: перцепція (ідентифікація явищ у відео), аналіз (транскрибування та індуктивне відкриття закономірностей за принципом DDL – навчання на основі даних) та продукція (імітація, самозапис). Студенти діяли як «лінгвістичні детективи», самостійно досліджуючи реалізації звуків у різних акцентах і контекстах, що сприяло розвитку їхньої автономії.

Результати дослідження продемонстрували статистично значущий прогрес. Зокрема, ідентифікація зв'язного мовлення зросла на 34%, а навички редукції голосних у слабких формах покращилися на 42%. Отримані дані підтверджують, що поєднання візуальних артикуляційних підказок із корпусним аналізом є більш ефективним, ніж використання традиційних дедуктивних фонетичних вправ. Дослідження доводить доцільність упровадження автентичних відеокорпусів у курс практичної фонетики англійської мови для розвитку навичок аудіювання і продуктивного мовлення, а також забезпечення високого рівня адаптивності до реальних ситуацій усної іншомовної комунікації з урахуванням варіативності дискурсів і акцентного розмаїття англійської мови.

Ключові слова: *фонетична компетентність, зв'язне мовлення, відеокорпуси, навчання на основі даних (DDL), мультимодальне навчання, англійська як мова міжнародного спілкування, автономія здобувачів.*

Received: March 01, 2026
Accepted: March 23, 2026
Published: April 22, 2026