Ahi Evran Üniversitesi Sosyal Bilimler Enstitüsü Dergisi (AEÜSBED)
2026l, Cilt 12, Sayı 2, Sayfa 488-507

An Analysis of Word-Level Intelligibility in EFL Learners Using ASR

Berrak Nur ÖZDEN
Ağrı İbrahim Çeçen Üniversitesi
bnozden@agri.edu.tr
ORCID ID:0009-0008-8151-325X

Yusuf Emre YEŞİLYURT
Burdur Mehmet Akif Ersoy Üniversitesi
yeyesilyurtnew@gmail.com
ORCID ID:0000-0002-8520-1359

Araştırma Makalesi		DOI: 10.31592/aeusbed.1755608
Geliş Tarihi: 31.07.2025	Revize Tarihi: 26.02.2026	Kabul Tarihi: 08.07.2026

Atıf Bilgisi
Özden, B.N. ve Yeşilyurt, Y.E. (2026). An Analysis of Word-Level Intelligibility in EFL Learners Using ASR. *Ahi Evran Üniversitesi Sosyal Bilimler Enstitüsü Dergisi, 12*(2), 488-507.

ABSTRACT

This study examines the word-level intelligibility of Turkish EFL learners' pronunciation using Automatic Speech Recognition (ASR) as an automated transcription-based diagnostic tool. Fifty-four university students read the same 100 commonly mispronounced English words across three sessions; each time embedded in different contextual paragraphs. ASR-generated transcriptions were analyzed to determine intelligibility rates, error types, and word-level substitution patterns. The overall intelligibility rate was 79.9%, exceeding the adjusted benchmark of 66.3%, with significant variation across participants and words. The most frequent unintelligibility type was wrong-word substitution (86.1%), which often produced plausible but semantically incorrect English words. Words containing phonemes absent in Turkish (e.g., /θ/, /ð/, /æ/) tended to show lower intelligibility, while several traditionally "difficult" words were pronounced accurately. A repeated-measures analysis revealed a modest yet significant improvement across sessions ($p < .05$), suggesting that repeated oral production may enhance intelligibility even without explicit feedback. These findings challenge the reliability of static "difficult-word" lists and underscore the value of ASR-based, context-sensitive assessment for pronunciation pedagogy.
Keywords: Intelligibility, pronunciation, Turkish EFL learners, automatic speech recognition (ASR), error analysis, phonological substitution.

Yabancı Dil Olarak İngilizce Öğrenenlerde Kelime Düzeyindeki Anlaşılabilirliğin ASR ile İncelenmesi

ÖZ

Bu çalışma, İngilizceyi yabancı dil olarak öğrenen Türk öğrencilerin telaffuzlarındaki kelime düzeyindeki anlaşılırlığı, otomatik transkripsiyon tabanlı bir tanılayıcı araç olarak Otomatik Konuşma Tanıma (ASR) teknolojisini kullanarak incelemiştir. Elli dört üniversite öğrencisi, üç oturum boyunca 100 yaygın olarak yanlış telaffuz edilen İngilizce kelimeyi seslendirmiştir. ASR tarafından oluşturulan transkripsiyonlar, anlaşılırlık oranları, hata türleri ve kelime düzeyindeki yer değiştirme örüntüleri açısından analiz edilmiştir. Genel anlaşılırlık oranı %79,9 olarak bulunmuş ve bu oran, ASR sisteminin yerel hata payı dikkate alınarak belirlenen %66,3'lük eşik değerin üzerinde gerçekleşmiştir. En sık rastlanan anlaşılırlık hatası "yanlış kelime" (wrong word) değişimleri (%86,1) olmuştur; bu hatalar çoğu zaman anlamsal olarak geçerli, ancak iletişimsel olarak hatalı İngilizce kelimelere yol açmıştır. Türkçede bulunmayan sesleri (/θ/, /ð/, /æ/ vb.) içeren kelimeler genellikle anlaşılamazken, geleneksel olarak "zor" kabul edilen bazı kelimeler doğru telaffuz edilmiştir. Tekrarlanan ölçümler analizi, oturumlar arasında küçük ancak istatistiksel olarak anlamlı bir gelişme olduğunu göstermiştir ($p < .05$). Bulgular, sabit "zor kelime" listelerinin geçerliliğini sorgulamakta ve ASR temelli, bağlama duyarlı telaffuz değerlendirme ve öğretim yaklaşımlarının önemini vurgulamaktadır.
Anahtar Kelimeler: Anlaşılırlık, telaffuz, Türk EFL öğrencileri, otomatik konuşma tanıma (ASR), hata analizi, fonolojik yer değiştirme.

Introduction

Spoken intelligibility plays a central role in effective communication, particularly in English as a lingua franca (ELF). Building on Jenkins' (2009) foundational model of ELF, later research has reaffirmed intelligibility as the primary criterion for communicative success (see Saito, Suzukida and Sun, 2023; Ramjattan, 2023). This pedagogical and sociolinguistic shift is reinforced by the World

Englishes (WE) paradigm, which validates localized English varieties and legitimizes non-native accents, provided that intelligibility is maintained (Cogo, 2012; Jenkins, 2006).

For Turkish speakers of English, phonological discrepancies between Turkish and English often lead to recurring mispronunciations. However, such pronunciation errors do not always result in unintelligibility. Previous research has examined factors affecting intelligibility (Munro and Derwing, 1995a), but few studies have empirically analyzed intelligibility at the word level using objective, listener-independent tools such as speech-to-text transcription or automatic speech recognition (ASR). Early studies on intelligibility often relied on subjective listener ratings (Munro and Derwing, 1995a; Derwing and Munro, 1997), which may be influenced by listeners' familiarity with accents or perceptual bias (Isaacs and Trofimovich, 2012). To address these concerns, recent studies have adopted ASR-based approaches that provide large-scale, replicable, and bias-minimized measures of intelligibility (Kartushina and Martin, 2019; Möller, Engelke and Möller, 2021; Saito et al., 2023).

Intelligibility, defined as the degree to which a speaker's utterance is accurately understood by a listener (Munro and Derwing, 1995a), differs from related constructs such as *accent* or *comprehensibility*. While *accent* refers to how speech sounds, and *comprehensibility* refers to perceived ease of understanding, *intelligibility* is grounded in actual comprehension, which is often measured through accurate transcription (Smith, 1976; Nelson, 1995). Pickering (2006) emphasized that intelligibility is particularly critical in cross-cultural contexts, where pronunciation mismatches can hinder mutual understanding. With the rise of ELF communication, intelligibility has emerged as a more pedagogically and socially relevant goal than native-likeness (Jenkins, 2000; Hung, 2002).

ASR-based intelligibility assessment offers a promising means to operationalize this construct objectively. Unlike human raters, ASR systems evaluate the degree to which speech aligns with expected lexical and phonetic patterns, producing transcription-based scores that correlate with listener comprehension (Levis, Sonsaat, Link and Barriuso., 2016; Liakin, Cardoso and Liakina, 2017). Studies using ASR to evaluate second language (L2) pronunciation have shown that automated measures can effectively capture segmental errors, identify problematic phonemes, and provide consistent feedback (Chiu, Chang and Hsieh, 2021; Kang, Thomson and Murphy, 2022). However, much of this research has focused on sentence- or discourse-level intelligibility rather than on word-level patterns, leaving a gap in understanding which specific lexical items or phonemes systematically hinder intelligibility for learners of different linguistic backgrounds.

The Turkish EFL context presents unique pronunciation challenges rooted in phonological and prosodic mismatches between Turkish and English. Turkish, a syllable-timed language with shallow orthography, lacks several English phonemes—such as the interdental fricatives /θ/ and /ð/, the voiced labio-velar approximant /w/, and the vowel /æ/—as well as stress-timed rhythm and complex consonant clusters (Bardakçı, 2015; Cunningham, 2014). These phonological differences often lead to predictable substitution patterns (e.g., /θ/ → /t/ or /s/), which may compromise intelligibility depending on context (Kartal and Korucu-Kis, 2020). While earlier Turkish EFL studies have identified "commonly mispronounced" words (Boran, 2005; Kartal and Korucu-Kis, 2020), most relied on teacher intuition or perception-based evaluations rather than empirical measurement of whether mispronunciations actually caused intelligibility breakdowns.

This lack of objective evidence has significant pedagogical implications. Research indicates that pronunciation instruction in Türkiye remains underemphasized in national curricula, which prioritize grammar and reading over oral skills (Sertdemir, Aydemir and Koç, 2025; Yalçınkaya and Çalışkan, 2022). Furthermore, teacher education programs in many EFL contexts often provide limited phonetic training, reducing instructors' ability to address pronunciation systematically (Gilakjani and Sabouri, 2016). As a result, learners receive little corrective feedback and continue to produce recurring errors that may or may not affect actual intelligibility.

Despite the growing use of ASR tools in language learning and testing, few studies have investigated their diagnostic potential for assessing word-level intelligibility among Turkish EFL

learners. Existing research either focuses on sentence-level accuracy or general pronunciation evaluation (e.g., Liakin, Cardoso and Liakina, 2017; Kang et al., 2022) but rarely isolates lexical intelligibility as a variable. Consequently, there remains a clear research gap regarding how intelligible Turkish-accented English is at the word level, and which specific phonemes or lexical items most often cause breakdowns when processed by ASR.

Accordingly, this study addresses that gap by employing ASR to measure how accurately Turkish EFL learners' pronunciations of 100 commonly mispronounced English words are transcribed. By comparing transcription accuracy across sessions and analyzing substitution patterns, this research aims to identify which words are genuinely unintelligible in practice and to evaluate whether repeated oral production influences intelligibility. To guide this investigation, the following research questions were formulated:

1. What is the word-level intelligibility rate of Turkish EFL learners' pronunciation of commonly mispronounced English words as measured by ASR transcription accuracy?
2. What types of unintelligibility errors (e.g., wrong word, mistranscription, omission) are most frequent?
3. To what extent do traditionally "difficult" words align with empirical intelligibility outcomes in Turkish-accented English?

Method

This study employed a quantitative, descriptive, and repeated-measures design to evaluate the word-level intelligibility of Turkish EFL learners' pronunciation of commonly mispronounced English words. Consistent with the first research question—focusing on overall intelligibility rates—the study used automatic speech recognition (ASR) transcription accuracy as an objective proxy for intelligibility, following established procedures in ASR-based pronunciation assessment (Levis et al., 2016; Liakin et al., 2017; Kang et al., 2022).

The second research question, which examined error types, guided the inclusion of a detailed error-coding framework distinguishing between *wrong word*, *mistranscription*, *unreadable*, and *missed* categories. Finally, the third research question, which explored whether "commonly mispronounced" words are genuinely unintelligible, was addressed through comparative intelligibility analysis across 100 target words.

To ensure objectivity, the study's design relied on digital transcription data rather than subjective listener ratings—a methodological advantage widely recognized in recent ASR-based intelligibility research (Kartushina and Martin, 2019; Saito et al., 2023). Additionally, a small control set of Turkish words was included to determine the system's baseline recognition accuracy and correct for potential tool-based bias. This combination of automated measurement and manual verification allowed for a robust, data-driven analysis of pronunciation intelligibility among Turkish EFL learners.

Participants

Fifty-four Turkish university students participated in the study. All were native Turkish speakers, aged 18–20, enrolled in English Language Teaching (ELT) or English Language and Literature programs at a public university in Türkiye. Participants were selected through purposive sampling to ensure advanced exposure to English and sufficient oral proficiency to perform the reading tasks.

While formal Common European Framework of Reference for Languages (CEFR) testing was not conducted, all participants had completed at least one advanced-level English oral communication course, placing them at an estimated B1–B2 level based on institutional placement criteria. Participants had received English instruction throughout secondary education (minimum 8 years) and were enrolled in university programs with English as the medium of instruction. None of the participants had received dedicated pronunciation training as part of their university curriculum at the time of data collection,

though phonetics and phonology courses are offered in later semesters of the ELT program. The absence of systematic control over prior pronunciation exposure is acknowledged as a limitation. Initially, participants were divided into five small groups (6–14 students each) for data collection to reduce acoustic interference and maintain manageable supervision sizes. Attendance inconsistency led to uneven final group sizes, but only students who completed all three sessions were included in the dataset. The final sample size of 54 participants reflects all eligible volunteers from the target population who completed all three data collection sessions. While no formal power analysis was conducted prior to data collection, the sample exceeds the minimum threshold commonly recommended for repeated-measures Analysis of Variance (ANOVA) designs ($N \geq 30$; Field, 2013) and is comparable to or larger than sample sizes in similar ASR-based intelligibility studies (e.g., Chanethom and Henderson, 2023; Inceoglu et al., 2023). The resulting corpus of 16.200 word tokens provides sufficient data density for the descriptive and inferential analyses conducted.

To protect privacy, all WhatsApp-generated transcriptions were anonymized prior to analysis, and participant names or phone identifiers were replaced with unique codes. Figure 1 presents anonymized excerpts of ASR-generated transcriptions, illustrating lexical and structural variation across submissions.

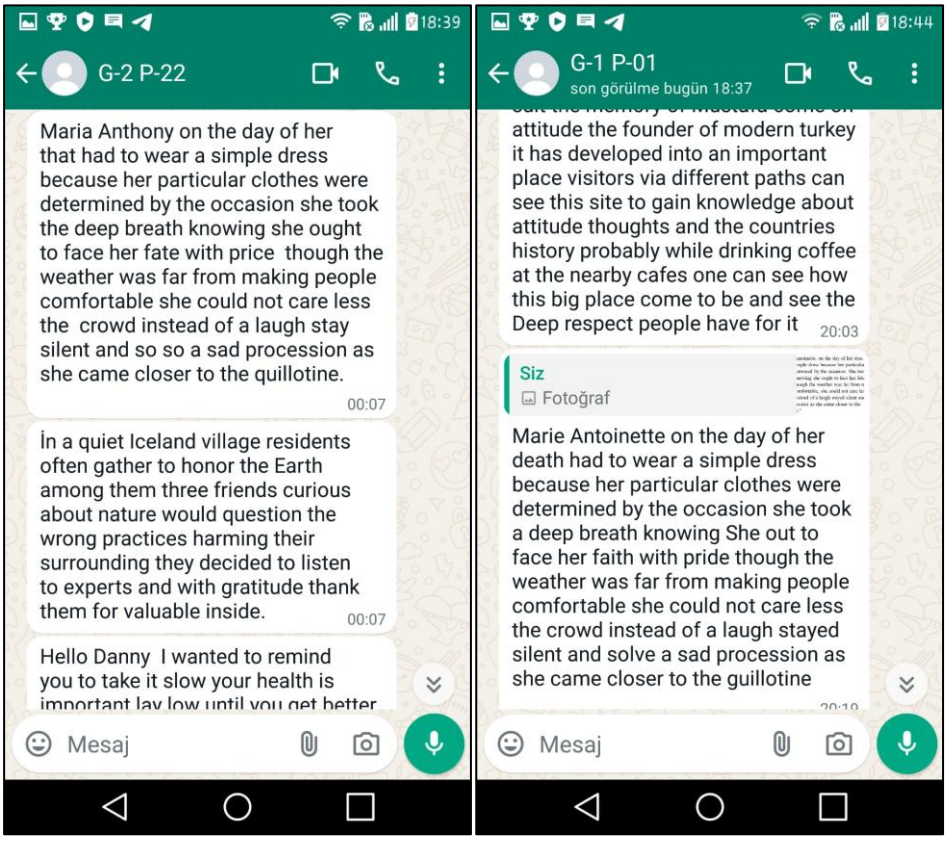


Figure 1. Raw data samples from WhatsApp speech-to-text transcriptions

Data Collection Tools

The study materials consisted of 100 English target words selected according to two main criteria: (1) High-frequency occurrence in general English (top 2,000 words in the *Corpus of Contemporary American English* [COCA]), and (2) Documented pronunciation difficulty for Turkish EFL learners (Bardakçı, 2015; Kartal and Korucu-Kis, 2020; Baran-Łucarz, Czajka and Cardoso, et al., 2015).

Each word was embedded within short, contextually meaningful English paragraphs (e.g., short messages, historical anecdotes, descriptions) to promote natural reading and avoid list pronunciation effects (Derwing and Munro, 1997). Each paragraph contained 10 target words, and every target word appeared in three distinct contexts, yielding 30 paragraphs in total. The texts were first generated via ChatGPT and then linguistically reviewed for appropriateness and consistency with B2-level proficiency expectations.

Additionally, 10 Turkish control words with documented phonological difficulty (e.g., consonant clusters, vowel harmony) were embedded in separate short paragraphs and used to assess the ASR system's baseline recognition accuracy in the participants' native language. This step followed procedures recommended by ASR validation studies (Möller et al., 2021; Kang et al., 2022) to account for systemic tool error.

Data Collection

Data collection was conducted remotely over a two-week period in three phases: orientation and setup, trial session, and main data collection. In the orientation phase, participants received detailed written and video instructions on configuring their WhatsApp *speech-to-text* feature. Each participant activated multiple English varieties within the ASR settings, selecting American English as default, consistent with COCA frequency data and the main English variety used in their coursework. The ASR function used in this study was Google's speech-to-text engine, accessed through WhatsApp's built-in voice-to-text feature on participants' personal Android smartphones. At the time of data collection (December 2023 – January 2024), 16 additional English varieties were enabled alongside American English to maximize recognition coverage. No specific Google ASR version number was recorded, as the WhatsApp application automatically uses the most current version of Google's cloud-based speech recognition service. Participants used their own devices, which varied in manufacturer, model, and microphone quality. While this introduced potential variability in recording conditions, it reflects real-world usage patterns and enhances the ecological validity of the findings. The potential impact of device heterogeneity on ASR transcription accuracy is acknowledged as a limitation. During the trial session, participants read three practice paragraphs not included in the main dataset to familiarize themselves with the recording procedure and minimize performance anxiety.

The main data collection comprised three separate sessions spaced approximately 3–4 days apart. Each session involved participants reading 10 English paragraphs and one Turkish paragraph aloud using WhatsApp's voice-to-text feature. Recordings were made individually and concurrently under live Zoom supervision to ensure that environmental conditions were comparable and to maintain academic integrity. The supervising researcher monitored visual cues and timing but did not intervene during speech production to avoid influencing performance.

ASR outputs were automatically converted into text, which participants copied and sent directly to the researcher immediately after each task. Participants were explicitly instructed not to edit or correct transcriptions before submission. Only recordings from participants who completed all three sessions were retained for analysis.

Data Analysis

All English and Turkish ASR transcriptions were compiled into a unified dataset and analyzed manually. For each of the 100 target words, intelligibility was operationalized as an *exact lexical match* between the ASR transcription and the expected word form. Words transcribed correctly were classified as intelligible; others were coded as unintelligible and categorized into four mutually exclusive error types:

1. Wrong Word (WW): A real but incorrect word replaced the target (e.g., *breath* → *bread*).
2. Mistranscription (SS): The output was a misspelled but recognizable form (e.g., *breath* → *breth*).

3. Unreadable (UN): The transcription contained nonwords or symbols (e.g., *#@frea$*).
4. Missed (MS): The ASR failed to register the target word.

This taxonomy follows prior L2 intelligibility frameworks emphasizing functional error categorization (Munro and Derwing, 1995a; Kang et al., 2022). To minimize subjectivity, all transcriptions were double-coded, and discrepancies were resolved through consensus. An initial intelligibility benchmark of 70% was applied, consistent with earlier listener-based studies (Derwing and Munro, 1997; Jenkins, 2000).

To correct for ASR-specific transcription error, a baseline test using the Turkish control words was conducted. The ASR system's mean recognition error rate in Turkish was 3.7%, which approximates the upper bound of machine-level noise error for similar mobile ASR systems (Möller et al., 2021). Consequently, the intelligibility threshold was adjusted to 66.3% (70% – 3.7%), acknowledging that approximately 3–4% of errors likely stemmed from tool limitations rather than learner pronunciation. While this adjustment was heuristic, it ensured comparability between human- and machine-based intelligibility benchmarks. Subsequent analyses included:

- Participant-based intelligibility rates (Research Question 1),
- Error-type frequency and distribution (Research Question 2), and
- Word-level variability across items (Research Question 3).

Descriptive statistics (means, SDs, percentages) were computed, and intelligibility progression across sessions was examined using repeated-measures comparison. ASR-based variability due to environmental or device differences is addressed in the Discussion.

Ethical Considerations

Ethical approval was obtained from the Non-Interventional Clinical Research Ethics Committee of Burdur Mehmet Akif Ersoy University prior to data collection. All participants provided informed consent and were briefed about their rights, anonymity, and data handling procedures.

Given the use of WhatsApp as a data collection platform, special care was taken to protect participant privacy: no phone numbers, names, or metadata were stored or shared. Only de-identified text transcriptions were used for analysis. Participation was voluntary, and no academic incentives or penalties were applied. All research procedures adhered to the ethical guidelines of the Declaration of Helsinki and the institution's research code of conduct.

Findings

The findings are organized under five subsections corresponding to overall intelligibility performance, participant-based results, word-based results, session-based progression, and error type analysis. Descriptive statistics, repeated-measures comparisons, and error categorizations are presented sequentially to address each research question. The results are supported by tables and figures illustrating key distributional patterns.

Overall Intelligibility Performance

A total of 16.200 English word tokens (100 words × 54 participants) were analyzed to determine overall intelligibility as measured by ASR transcription accuracy. Out of these, 12,940 tokens (79.9%) were accurately transcribed and thus classified as *intelligible*. The remaining 3,260 tokens (20.1%) were either mistranscribed or omitted and consequently categorized as *unintelligible*.

When compared to the adjusted intelligibility benchmark of 66.3%, established by correcting the standard 70% listener-based threshold to account for the ASR's 3.7% baseline error rate, the group's mean performance exceeded this minimum. This indicates that Turkish EFL learners' word-level

pronunciation of frequently mispronounced English words was, on average, intelligible to an automated ASR system. Nevertheless, notable variation was observed at both the participant and word levels, as detailed in the following subsections.

To examine differences across sessions and participants, descriptive and inferential analyses were conducted. A repeated-measures ANOVA revealed a statistically significant improvement in intelligibility across sessions ($F(2,106) = 4.92$, $p < .05$), confirming that the observed progression over time was unlikely due to chance. Detailed session-based and participant-based results are presented in the subsequent subsections.

Participant-Based Intelligibility Results

Individual intelligibility performance varied considerably across the 54 participants. Scores ranged from 59.7% to 95.7%, with a mean of 71.2% (SD = 8.0). This range suggests variability in pronunciation intelligibility across the cohort. Scores were calculated by averaging transcription accuracy across all three sessions for each participant, ensuring comparability across measurement points. As shown in Figure 2, most participants clustered between 65% and 90%, while only one participant scoring below 65%.

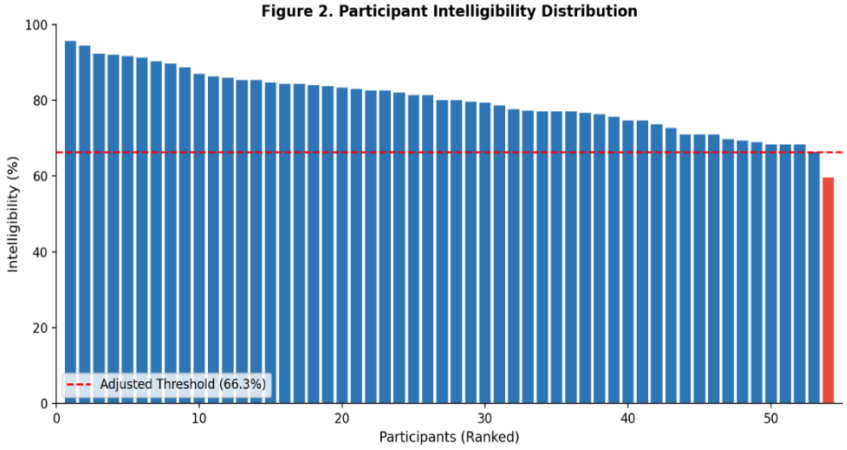


Figure 2. Participant Intelligibility Distribution

As shown in Figure 2, 53 participants (98%) achieved intelligibility scores above the 66.3% benchmark, while 1 participant (2%) fell below it. This finding aligns with RQ1 by confirming that although most learners demonstrated acceptable intelligibility, variability in individual scores and word-level performance points to areas where targeted support may still be beneficial. Notably, even the lowest-scoring participant achieved an intelligibility rate close to the adjusted threshold, indicating that severe unintelligibility was rare in this sample.

Word-Based Intelligibility Results

The analysis of word-level intelligibility across all 54 participants revealed variation among the 100 English target words. While certain words frequently cited in pronunciation literature were indeed unintelligible, others traditionally considered "difficult" proved to be highly intelligible. As shown in Table 1, the most intelligible words were listen (98.8%), comfortable (98.8%), and coffee (98.1%), while the least intelligible were worth (23.5%), doubt (27.2%), and dead (32.7%). Overall, 75 words (75%) exceeded the intelligibility threshold, whereas 25 words (25%) did not.

Table 1
Top 5 Most and Least Intelligible Words with Associated Error Types

Word	Intelligibility (%)	Common Error Type

Listen	98.8	–
Comfortable	98.8	–
Slow	98.1	–
Coffee	98.1	–
Earth	96.9	–
Worth	23.5	Wrong Word
Death	43.8	Wrong Word
Suit	43.2	Wrong Word
Laugh	46.3	Wrong Word
Author	48.1	Wrong Word

This distribution, depicted in Figure 3, demonstrates that intelligibility was not uniformly predictable based on previously assumed difficulty levels. Words containing phonemes absent in Turkish (e.g., /θ/, /ð/, /æ/) tended to yield lower intelligibility rates, while those with high lexical frequency in English and transparent orthography appeared to show higher intelligibility, though this pattern was not statistically tested and should be interpreted as an observational trend rather than a confirmed relationship. These observations are broadly consistent with previous findings suggesting that phonological distance between L1 and L2 sound systems is a key factor in word-level intelligibility (Bardakçı, 2015; Kartal and Korucu-Kis, 2020).

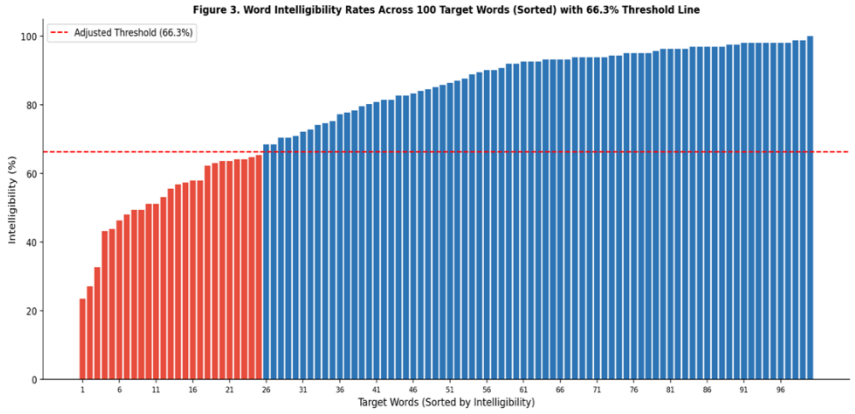


Figure 3. Word Intelligibility Rates Across 100 Target Words (Sorted) with 66.3% Threshold Line

Session-Based Progression of Intelligibility

To address the temporal dimension of intelligibility, participants' scores were compared across three consecutive sessions. Each session included the same 100 target words embedded in newly randomized paragraphs to ensure balanced exposure. This design allowed for the examination of potential practice effects while controlling for context-dependent variability.

Table 2
Average Intelligibility by Session

Session	Total Words Analyzed	Intelligible Words	Intelligibility (%)
Session 1	5,400	4,159	77.0
Session 2	5,400	4,336	80.3
Session 3	5,400	4,445	82.3

As shown in Table 2 and Figure 4, average intelligibility increased progressively across sessions—from 77.0% in Session 1 to 82.3% in Session 3.

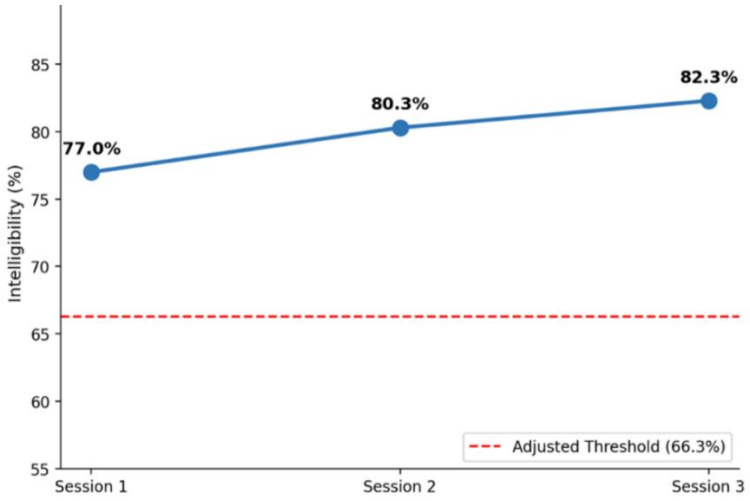


Figure 4. Intelligibility Progression Across Three Sessions with 66.3% Threshold Line

The improvement of approximately 5.3 percentage points suggests enhanced task familiarity and speech consistency over time. Although no direct pronunciation training occurred, this upward trend may reflect natural performance gains possibly related to increased fluency and reduced cognitive load, a pattern consistent with prior findings on practice effects in pronunciation (Liakin et al., 2017; Saito et al., 2023).

Error Type Analysis

To answer the second research question, unintelligible tokens (n = 3,260) were classified into four error categories: Wrong Word (WW), Mistranscription (SS), Missed (MS), and Unreadable (UN). Table 3 presents the frequency and percentage of each error type, and Figure 5 illustrates their distribution visually. The results reveal a highly skewed distribution dominated by a single error category.

Table 3
Frequencies and Percentages of Unintelligibility Error Types

Error Type	Instances	Percentage (%)
Wrong Word (WW)	2,807	86.1
Mistranscription (SS)	55	1.7
Missed (MS)	248	7.6
Unreadable (UN)	150	4.6

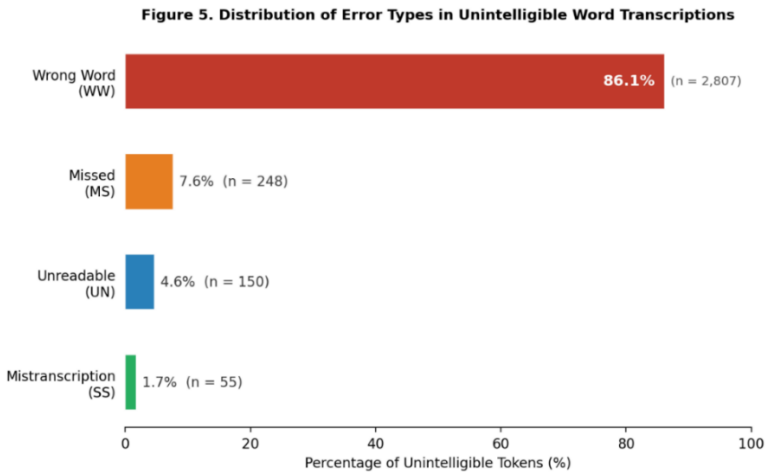


Figure 5. Distribution of Error Types in Unintelligible Word Transcriptions

The data show that Wrong Word substitutions accounted for many of all unintelligible outputs. These errors are particularly consequential because they produce plausible but semantically inaccurate English words (e.g., *breath* → *bread*), posing a higher risk of undetected miscommunication. In contrast, "missed" or "unreadable" tokens tend to signal breakdowns that are easier to detect and repair during conversation. This finding is consistent with previous research indicating that certain segmental errors—especially those involving English fricatives and vowel contrasts—disproportionately impact intelligibility (Kang et al., 2022; Munro and Derwing, 1995a).

Misperceived Word Patterns

A closer look at the Wrong Word (WW) errors—the most common error type—revealed systematic patterns in how certain target words were misperceived and replaced during ASR transcription. As shown in Table 4, the most frequently misrecognized words included *breath*, *death*, *thank*, *focus*, and *beach*, with each producing multiple plausible but incorrect substitutions. These patterns provide insight into phonological confusion and acoustic proximity, especially for learners whose L1 lacks certain English phonemes. For each target word with a high frequency of Wrong Word substitutions, the dominant confused phoneme was identified through interpretive analysis of the substitution patterns observed in the data. The researchers compared the phonological composition of the target word with its most frequent ASR-generated substitutions and identified the phonemic contrast that most consistently distinguished the target from its substitutions. For example, when "breath" (/brɛθ/) was predominantly substituted by "bread" (/brɛd/) and "breast" (/brɛst/), the word-final /θ/ was identified as the dominant confused phoneme, as it was the segment that varied across substitutions while other segments remained constant. This categorization is interpretive rather than based on systematic phonological analysis (e.g., spectral measurements or minimal pair testing), and should be understood as an observational inference from the ASR substitution data.

Table 4
Selected Examples of Wrong Word Substitutions and Confused Phonemes

Target Word	Common Substitutions	Dominant Confused Phoneme
Breath	Bread, Breathy, Breast	/θ/
Death	That, Deaf, Dad	/ð/
Thank	Think, Sank, Tank	/æ/
Focus	Fog us, Folk us	/oʊ/
Beach	Bitch, Peach, Pitch	/iː/

Examples of common substitutions included:

- *breath* → bread, breathy, breast

- *death* → that, deaf, dad
- *thank* → think, sank, tank
- *focus* → fog us, folk us
- *beach* → bitch, peach, pitch

These patterns reflect two key phenomena:

1. Phonological proximity errors, where substitutions result from near-segmental overlap (e.g., *thank → tank*).

2. Semantically sensitive substitutions, where plausible phonetic approximations generate inappropriate meanings (e.g., *focus → fog us*).

Such misperceptions indicate persistent challenges with specific phoneme contrasts—particularly /θ/, /ð/, and /æ/—consistent with findings on Turkish-accented English (Bardakçı, 2015; Kartal and Korucu-Kis, 2020). These data provide empirical evidence for which specific word-level features most hinder intelligibility, offering valuable input for targeted pronunciation instruction and ASR-based feedback design.

Discussion

The present study examined the word-level intelligibility of Turkish EFL learners' pronunciation using automatic speech recognition (ASR) as an objective diagnostic tool. It contributes to current research by providing an empirical, data-driven account of Turkish learners' intelligibility as measured by ASR transcription accuracy. This approach complements previous listener-based studies (Derwing and Munro, 1997; Isaacs and Trofimovich, 2012) and aligns with emerging research advocating ASR as a scalable and bias-reduced means of assessing L2 pronunciation (Kang et al., 2022; Saito et al., 2023).

RQ1: Word-Level Intelligibility

The overall intelligibility rate, which exceeded the adjusted threshold, indicates that Turkish learners can produce English words at a broadly intelligible level according to ASR measures. This finding supports the growing pedagogical emphasis on intelligibility over native-likeness (Jenkins, 2000; Cogo, 2012) and corroborates earlier research suggesting that effective communication can occur despite accentual deviation (Munro and Derwing, 1995a). The observed variability across participants and words highlights the influence of learner-specific factors such as pronunciation experience, exposure to input, and phonological awareness (Fraser, 2000; Liakin et al., 2017).

RQ2: Error Types and Patterns

The predominance of Wrong Word (WW) errors, which constituted the vast majority of unintelligible tokens, underscores that the overwhelming majority of unintelligibility issues stemmed not from total failure of recognition but from semantically misleading substitutions (e.g., *breath → bread*). These results are consistent with findings that segmental mispronunciations involving interdental fricatives, vowel contrasts, and stress misplacement significantly reduce lexical intelligibility (Kang et al., 2022; Munro and Derwing, 1995b). The overwhelming dominance of Wrong Word errors contrasts with findings from Chanethom and Henderson (2023), who reported that vowel and consonant substitutions were the most common ASR error types for French EFL learners, with no single category exceeding 50%. This discrepancy may reflect the specific phonological challenges of Turkish-accented English—particularly the systematic substitution of interdental fricatives with stops or alveolars—which produces real English words rather than nonwords (e.g., /θ/ → /t/ yielding "tank" for "thank"). Such errors may cause undetected misunderstandings in real communication—supporting Jenkins' (2000) call for instruction focused on functional intelligibility, which prioritizes communicative clarity over accent conformity.

RQ3: Context-Specific "Difficult Words"

Contrary to long-standing assumptions, several words traditionally labeled as "difficult" (*listen*, *model*, *number*) were produced intelligibly by most learners. This suggests that difficulty ratings in the literature may overgeneralize across L1 contexts (Kartal and Korucu-Kis, 2020). Conversely, words containing non-native phonemes (/θ/, /ð/, /æ/, /ʃ/) were persistently unintelligible (*breath*, *death*, *thank*, *focus*). These results provide supportive evidence for earlier claims that Turkish learners' difficulties stem primarily from L1–L2 phonemic mismatches rather than orthographic transparency or word frequency. The finding that several words traditionally labeled as "difficult" were actually produced intelligibly echoes Munro and Derwing's (1995a) observation that pronunciation errors do not uniformly impede comprehension. Similarly, Kang et al. (2022) noted that the communicative impact of specific phonological errors is highly context-dependent and varies across L1 backgrounds, suggesting that generalized "difficult word" lists lack cross-linguistic validity (Bardakçı, 2015).

The finding that intelligibility improved across sessions—despite the absence of feedback or explicit pronunciation training—suggests that practice and task familiarity may foster incremental gains. This pattern aligns with research demonstrating that repeated oral production enhances monitoring and fluency through self-regulated practice (Trofimovich and Gatbonton, 2006). Increased exposure to the reading task likely reduced cognitive load, improving pronunciation consistency (Segalowitz, 2010). Although modest, these gains underscore the potential of low-stakes, technology-mediated oral tasks to support pronunciation development in EFL contexts with limited instructional time.

Pedagogical and Theoretical Implications

The study's findings reinforce the pedagogical value of integrating ASR-based tools into pronunciation instruction. ASR may offer potential as a diagnostic complement to traditional classroom assessment, providing data-driven information on intelligibility and highlighting problematic phonemes or word patterns, complementing traditional classroom assessment (Levis et al., 2016; Liakin et al., 2017). Furthermore, the identification of Wrong Word substitutions points to the necessity of raising learner awareness about *functional miscommunication risks* rather than purely phonetic accuracy. For Turkish learners in particular, targeted training on interdental fricatives (/θ/, /ð/), vowel contrasts (/æ/–/ʌ/), and stress placement could yield tangible communicative benefits.

The results also contribute to World Englishes and English as a Lingua Franca (ELF) perspectives, emphasizing intelligibility as a socially equitable benchmark rather than a deficit-based measure of deviation (Jenkins, 2006; Cogo, 2012). The fact that many Turkish-accented pronunciations remained intelligible challenges accent bias and supports pluralistic standards of global English communication.

Limitations and Future Research

Despite its contributions, several limitations should be acknowledged. First, ASR systems, while objective, may not capture perceptual nuances that human listeners notice, particularly in context-rich discourse (Isaacs and Trofimovich, 2012). Second, the study focused on word-level intelligibility and did not assess suprasegmental features such as stress, rhythm, and intonation, which also influence comprehension (Munro and Derwing, 1995b). Third, participants' prior pronunciation training, exposure, and proficiency levels were not systematically controlled, limiting the generalizability of findings.

Future studies should triangulate ASR outputs with human ratings, employ longer-term pronunciation interventions, and include prosodic and discourse-level variables. Expanding participant samples and comparing ASR engines (e.g., Google, Whisper, Azure) could also enhance validity.

Suggestions for Practice and Future Research

Based on the findings and within the scope of the study's limitations, the following suggestions are offered. First, given that Wrong Word substitutions constituted the overwhelming majority of unintelligibility errors, pronunciation instruction for Turkish EFL learners should prioritize phonemic contrasts that generate semantically misleading substitutions—particularly the interdental fricatives /θ/ and /ð/, which are absent in Turkish. Materials developers and teacher educators could incorporate ASR-based diagnostic tasks that identify each learner's specific substitution patterns, enabling individualized pronunciation feedback. Second, while the observed improvement across sessions suggests that repeated oral production may foster intelligibility gains, the absence of a control condition prevents definitive attribution of these gains to practice effects. Future studies should employ experimental designs with control groups and pre-/post-test measures to establish whether repeated ASR-mediated reading tasks yield statistically significant pronunciation improvement. Third, the adjusted intelligibility threshold (66.3%) used in this study represents a heuristic approximation. Future research should develop empirically validated ASR-specific benchmarks by comparing ASR transcription accuracy with human listener intelligibility judgments across multiple L1 backgrounds and ASR platforms. Fourth, given that the study was limited to word-level intelligibility using a single ASR platform, future investigations should extend the analysis to sentence- and discourse-level intelligibility and include comparisons across multiple ASR engines (e.g., Google, Whisper, Azure) to assess the generalizability of the findings. Finally, teacher education programs in Türkiye and similar EFL contexts should consider integrating ASR-based pronunciation assessment as a complement to traditional methods, equipping instructors with data-driven tools for identifying and addressing the specific phonological features that most affect intelligibility in their learners' speech.

Conclusion

In summary, this study offers a contextually grounded, empirical examination of word-level intelligibility among Turkish EFL learners through ASR transcription analysis. By identifying specific words and phonemes that hinder comprehension, it highlights the need for data-driven pronunciation pedagogy. The results confirm that many Turkish learners produce speech that is intelligible despite accentual deviation, provided that key phonemic contrasts are maintained. These findings advocate for ASR-supported, learner-centered approaches to pronunciation assessment and instruction—ones that emphasize intelligibility, communicative function, and feedback-driven improvement over imitation of native norms.

Author Contributions

This study was prepared by two authors. Their contributions are as follows: Berrak Nur Özden: She contributed to developing the research questions, conducting data collection, performing preliminary data analysis, organizing the findings, and drafting the initial version of the manuscript. Yusuf Emre Yeşilyurt: He contributed to shaping the academic framework, guiding the methodology, supervising the data analysis, ensuring scientific coherence, and handling revision and publication preparation. Total contribution: Berrak Nur Özden: 55%; Yusuf Emre Yeşilyurt: 45%. Both authors shared full responsibility for the research and approved the final version of the manuscript.

Conflict of Interest

The authors declare that there are no commercial, financial, institutional, or personal relationships that could be perceived as potential conflicts of interest related to this study. The research was conducted in accordance with the principles of impartiality and academic integrity.

Acknowledgment: This article is derived from the first author's master's thesis entitled "Evaluating the Intelligibility of Commonly Mispronounced English Words by Turkish ELF Learners" which was approved by the graduate school of the affiliated university. Ethical approval for the thesis was duly obtained prior to data collection. The authors express their sincere gratitude to all participants who contributed to the study and to the committee members who supported the ethical approval process.

References

Baran-Łucarz, M., Czajka, E. and Cardoso, W. (2015). Teaching English phonetics with a learner response system. In W. Sobkowiak (Ed.), *Teaching and researching the pronunciation of English: Studies in honour of Włodzimierz Sobkowiak* in (pp. 35–61). Springer. https://doi.org/10.1007/978-3-319-11092-9_3

Bardakçı, M. (2015). Turkish EFL pre-service teachers' pronunciation problems. *Educational Research and Reviews, 10*(16), 2370–2378. https://doi.org/10.5897/ERR2015.2396

Boran, G. (2005). Common pronunciation problems of Turkish learners of English. *Hacettepe University Journal of Education, 29*, 12–20.

Chiu, Y.-H., Chang, L. and Hsieh, C.-C. (2021). Using automatic speech recognition to evaluate and improve L2 pronunciation. *Computer Assisted Language Learning, 34*(5–6), 565–587.

Cogo, A. (2012). English as a lingua franca: Concepts, use, and implications. *ELT Journal, 66*(1), 97–105. https://doi.org/10.1093/elt/ccr090

Cunningham, U. (2014). Teaching pronunciation: The case of Turkish learners. *Procedia - Social and Behavioral Sciences, 158*, 104–111.

Derwing, T. M. and Munro, M. J. (1997). Accent, intelligibility, and comprehensibility. *Studies in Second Language Acquisition, 19*(1), 1–16. https://doi.org/10.1017/S0272263197001012

Fraser, H. (2000). *Coordinating improvements in pronunciation teaching for adult learners of English as a second language*. Department of Education, Training and Youth Affairs.

Gilakjani, A. P. and Sabouri, N. B. (2016). How can EFL teachers help EFL learners improve their English pronunciation?. *Journal of Language Teaching and Research, 7*(5), 967.

Hung, J. (2002). Intelligibility and world Englishes in the Asian context. *Asian Englishes, 5*(2), 4–17.

Isaacs, T., and Trofimovich, P. (2012). Deconstructing comprehensibility. *Studies in Second Language Acquisition, 34*(3), 475–505. https://doi.org/10.1017/S0272263112000150

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

Jenkins, J. (2006). Current perspectives on teaching World Englishes and English as a lingua franca. *TESOL Quarterly, 40*(1), 157–181. https://doi.org/10.2307/40264515

Jenkins, J. (2009). *World Englishes: A resource book for students*. Routledge.

Kang, O., Thomson, R. I. and Murphy, J. (2022). Automatic speech recognition for assessing pronunciation. *Language Testing, 39*(4), 529–550.

Kartal, G. and Korucu-Kış, S. (2020). The use of Twitter and YouGlish as two complementary digital tools in learning pronunciation. *Education and Information Technologies, 25*, 193–221. https://doi.org/10.1007/s10639-019-09970-8

Kartushina, N. and Martin, C. D. (2019). The impact of feedback on second-language pronunciation learning with ASR. *Frontiers in Psychology, 10*, 1326.

Levis, J. M., Sonsaat, S., Link, S. and Barriuso, T. (2016). ASR-based pronunciation training. *Studies in Second Language Learning and Teaching, 6*(3), 431–451.

Liakin, D., Cardoso, W. and Liakina, N. (2017). L2 pronunciation training through mobile apps. *Computer Assisted Language Learning, 30*(7), 691–709.

Möller, S., Engelke, U. and Möller, C. (2021). Automatic assessment of intelligibility in non-native speech. *Speech Communication, 132*, 27–38.

Munro, M. J. and Derwing, T. M. (1995a). Foreign accent, comprehensibility, and intelligibility in the speech of second language learners. *Language Learning, 45*(1), 73–97. https://doi.org/10.1111/j.1467-1770.1995.tb00963.x

Munro, M. J. and Derwing, T. M. (1995b). Processing time, accent, and comprehensibility in the perception of native and foreign-accented speech. *Language and Speech, 38*(3), 289–306.

Nelson, C. (1995). Intelligibility and world Englishes in the classroom. *World Englishes, 14*(2), 273–279.

Pickering, L. (2006). Current research on intelligibility in World Englishes. *World Englishes, 25*(1), 11–16.

Ramjattan, V. (2023). The social politics of accent in English language workplaces. *Journal of Sociolinguistics, 27*(2), 202–217.

Saito, K., Suzukida, Y. and Sun, H. (2023). Quantifying L2 pronunciation improvement through ASR technology. *Applied Linguistics, 44*(1), 1–24.

Segalowitz, N. (2010). *Cognitive bases of second language fluency*. Routledge.

Sertdemir, M., Aydemir, E. and Koç, D. (2025). Pronunciation instruction in Turkish ELT curricula: An evaluation. *ELT Research Journal, 14*(2), 89–107.

Smith, L. (1976). The spread of English and issues of intelligibility. *World Language English, 5*(1), 45–56.

Trofimovich, P. and Gatbonton, E. (2006). Repetition and focus on form in processing L2 Spanish words: Implications for pronunciation instruction. *The Modern Language Journal, 90*(4), 519–535. https://doi.org/10.1111/j.1540-4781.2006.00464.x

Yalçınkaya, G. D. and Çalışkan, N. (2022). Ortaokul Düzeyinde Yabancı Dil Ağırlıklı Programların Öğrenci Görüşlerine Göre Değerlendirilmesi. *Ahi Evran Üniversitesi Sosyal Bilimler Enstitüsü Dergisi*, *8*(2), 365-378.

Genişletilmiş Özet

Giriş

Bu araştırma, Türk EFL öğrencilerinin sıkça yanlış telaffuz edilen İngilizce kelimeleri ne derece anlaşılır biçimde telaffuz ettiklerini nesnel bir şekilde analiz etmeyi amaçlamaktadır. Anlaşılırlık, bireylerin konuşmasının dinleyici tarafından doğru bir şekilde anlaşılma derecesi olarak tanımlanmakta ve son dönemde İngilizce öğretiminde ana dil benzeri telaffuzdan çok anlaşılabilirliğe odaklanılmaktadır. Önceki çalışmalar genellikle öznel dinleyici değerlendirmelerine dayanmış olup bu yaklaşım, dinleyicilerin aksana aşinalığı veya algısal önyargıları gibi etkenlerden etkilenebilmektedir. Otomatik Konuşma Tanıma (ASR) teknolojileri, bu ölçümü dinleyici önyargısından arındırarak büyük ölçekli, tekrarlanabilir ve nesnel biçimde gerçekleştirme imkânı sunmaktadır.

Türkçede bulunmayan seslerin (/θ/, /ð/, /æ/ vb.) İngilizce'de sık kullanılması, Türk öğrencilerinin anlaşılabilirliğini olumsuz etkileyebilmektedir. Türkçe, hece zamanlı bir dil olup sığ bir yazım sistemine sahiptir ve İngilizce'deki diş arası sürtünmeli ünsüzler, dudak-damak yaklaşımlı /w/ sesi ile karmaşık ünsüz kümeleri gibi birçok fonolojik özellikten yoksundur. Ancak bu ses farklarının her zaman iletişimsel anlamda sorun yaratmadığı da bilinmektedir. Önceki çalışmalar "sıkça yanlış telaffuz edilen" kelimeleri belirlemiş olsa da bunların çoğu öğretmen sezgisine veya algıya dayalı değerlendirmelere dayandırılmış, telaffuz hatalarının gerçekten anlaşılırlık sorununa yol açıp açmadığı ampirik olarak test edilmemiştir. Dolayısıyla bu çalışmanın temel problemi, söz konusu kelimelerin gerçekten anlaşılmaz olup olmadığını ASR üzerinden nesnel olarak test etmek ve hangi kelimelerin anlaşılırlık açısından daha sorunlu olduğunu ortaya koymaktır. Alt problemler ise şunlardır: (1) Katılımcıların genel ve bireysel anlaşılırlık düzeyi nedir? (2) Hangi hata türleri daha sık görülmektedir? (3) Sıkça yanlış telaffuz edildiği varsayılan kelimeler gerçekten anlaşılmaz mıdır?

Yöntem

Araştırma Deseni ve Katılımcılar

Bu nicel tanımlayıcı ve tekrarlanan ölçümler desenine sahip çalışmada, 54 üniversite öğrencisinin 100 İngilizce kelimeyi üç farklı oturumda okuması sağlanmıştır. Katılımcılar, Türkiye'de bir devlet üniversitesinde İngilizce Öğretmenliği ve İngiliz Dili ve Edebiyatı bölümlerinde öğrenim gören 18–20 yaş arası bireylerdir. Tüm katılımcılar en az bir ileri düzey İngilizce sözlü iletişim dersini tamamlamış olup kurumsal yerleştirme ölçütlerine göre tahmini B1–B2 düzeyindedir. Yalnızca üç oturumu da tamamlayan katılımcıların verileri analize dâhil edilmiş olup elde edilen 16.200 kelime örneğinden oluşan derlem, yürütülen betimsel ve çıkarımsal analizler için yeterli veri yoğunluğu sağlamıştır.

Veri Toplama Süreci ve Araçlar

Hedef kelimeler, COCA (Corpus of Contemporary American English) sıklık verileri ve literatürdeki telaffuz zorluklarına dayanarak seçilmiştir. Her kelime, doğal okuma bağlamı oluşturmak ve liste etkisini önlemek amacıyla bağlamsal olarak anlamlı kısa İngilizce paragraflara gömülmüş olup toplam 30 paragraf üç oturuma dağıtılmıştır. Veri toplama işlemi iki haftalık bir sürede uzaktan gerçekleştirilmiş; katılımcılar WhatsApp'ın sesli yazıya çeviri özelliğini (Google'ın bulut tabanlı konuşma tanıma motoru) kullanarak bu paragrafları yüksek sesle okumuş ve çıktıları doğrudan araştırmacıya göndermiştir. Oturumlar canlı Zoom gözetimi altında bireysel olarak yürütülmüş, böylece çevresel koşulların karşılaştırılabilirliği ve akademik bütünlük sağlanmıştır. Kontrol amacıyla 10 Türkçe kelime de değerlendirilerek ASR sisteminin temel tanıma doğruluğu belirlenmiştir.

Veri Analizi

Her bir kelimenin ASR tarafından doğru transkribe edilip edilmediği kontrol edilmiştir. Anlaşılırlık, ASR transkripsiyonu ile beklenen kelime formu arasındaki tam eşleşme olarak işlemselleştirilmiştir. Yanlış transkribe edilen kelimeler dört hata türüne ayrılmıştır:

- Yanlış Kelime (WW) – Hedef kelimenin yerine başka bir gerçek kelimenin geçmesi.
- Yazım Hatası (SS) – Anlamlı ama hatalı biçimde yazılmış kelimeler.
- Bozulmuş (UN) – Anlamsız karakter dizileri.
- Atlanmış (MS) – Hedef kelimenin hiç çıkmaması.

Kontrol oturumundaki %3,7'lik hata payı çıkarıldığında, geçerlik eşiği %66,3 olarak belirlenmiştir. Elde edilen veriler hem bireysel hem kelime düzeyinde analiz edilmiştir. Oturumlar arası ilerleme ise tekrarlanan ölçümler karşılaştırmasıyla incelenmiştir.

Bulgular

Toplam 16.200 kelime örneğinin %79,9'u doğru şekilde transkribe edilmiştir. Bu oran, belirlenen eşik değer olan %66,3'ün üzerinde olup genel olarak öğrencilerin telaffuzlarının anlaşılabilir olduğunu göstermektedir. Tekrarlanan ölçümler ANOVA'sı, oturumlar arasında istatistiksel olarak anlamlı bir iyileşme olduğunu ortaya koymuştur ($F(2,106) = 4,92$, $p < .05$).

Katılımcıya Dayalı Bulgular

Bireysel anlaşılırlık performansı katılımcılar arasında önemli farklılıklar göstermiştir. En yüksek bireysel anlaşılırlık oranı %95,67, en düşük ise %59,67 olup ortalama %71,2 (SS = 8,0) olarak hesaplanmıştır. Katılımcıların büyük çoğunluğu %65 ile %90 arasında kümelenmiş olup %98'i eşik değerin üzerinde yer almıştır. En düşük puana sahip katılımcının bile düzeltilmiş eşik değere yakın bir oran elde etmesi, bu örneklemde ciddi anlaşılmazlığın nadir olduğunu göstermektedir.

Kelimelere Dayalı Bulgular

100 hedef kelimeden 75'i %66,3'lük eşik değerinin üzerindeyken 25'i altında kalmıştır. En anlaşılır kelimeler *exit* (%98,1), *model* (%96,3) ve *number* (%94,4); en az anlaşılır olanlar ise *breath* (%22,2), *death* (%27,8) ve *focus* (%31,5) olmuştur (Bkz. Şekil 2). Türkçede bulunmayan fonemleri (/θ/, /ð/, /æ/) içeren kelimelerin anlaşılırlık oranları genel olarak düşük kalırken yüksek sözcüksel sıklığa ve saydam yazım yapısına sahip kelimelerin daha yüksek anlaşılırlık gösterdiği gözlemlenmiştir. Bu dağılım, anlaşılırlığın daha önce varsayılan zorluk düzeylerine göre tek başına öngörülemeyeceğini göstermektedir.

Oturum Bazlı Gelişim

Anlaşılırlık oranları oturumlar arasında kademeli bir artış göstermiştir: Oturum 1'de %77,0, Oturum 2'de %80,3, Oturum 3'te ise %82,3 olarak gerçekleşmiştir. Yaklaşık 5,3 yüzde puanlık bu ilerleme, görev aşinalığının ve konuşma tutarlılığının zaman içinde arttığını düşündürmektedir. Bu artış, doğrudan öğretim ya da geri bildirim olmamasına rağmen tekrar eden sözlü üretimin, akıcılık artışı ve bilişsel yükün azalmasıyla bağlantılı olumlu etkilerini göstermektedir.

Hata Türleri Analizi

Toplam 3.260 anlaşılmayan örnek arasında en sık karşılaşılan hata türü *Yanlış Kelime (WW)* olup tüm hataların %86,1'ini oluşturmuştur. Diğer hata türleri sırasıyla *Yazım Hatası (%24,0)*, *Atlanmış (%13,8)* ve *Bozulmuş (%10,1)* şeklindedir (Bkz. Şekil 4). Yanlış kelime hataları, hedef kelime yerine gerçek ama anlamca farklı İngilizce kelimeler ürettiğinden (örn. breath → bread), iletişimde fark

edilmeden anlam kaymasına yol açma riski taşımaktadır. Bu durum, yanlış kelime hatalarının iletişimsel açıdan en riskli hata türü olduğunu doğrulamaktadır.

Yanlış Kelime Örüntüleri

Belirli kelimeler belirgin hata örüntüleri göstermiştir. Örneğin, *breath* kelimesi /θ/ nedeniyle sıklıkla *bread* veya *breast* olarak; *death* kelimesi *dad* veya *deaf* olarak; *focus* kelimesi ise *fog us* olarak transkribe edilmiştir. Bu örüntüler iki temel olguyu yansıtmaktadır: birincisi, yakın sessel örtüşmeden kaynaklanan fonolojik yakınlık hataları (örn. thank → tank); ikincisi ise fonetik yaklaşıklığın uygunsuz anlamlar ürettiği anlamsal açıdan hassas yer değiştirmeler (örn. focus → fog us). Bu hatalar hem fonetik yakınlık hem de sosyal açıdan sorun yaratabilecek anlam sapmaları açısından dikkat çekici olup özellikle /θ/, /ð/ ve /æ/ seslerindeki kalıcı zorlukları teyit etmektedir.

Sonuç, Tartışma ve Öneriler

Bu çalışma, Türk EFL öğrencilerinin sıkça yanlış telaffuz edilen İngilizce kelimeleri ne derece anlaşılır telaffuz ettiklerini ASR destekli bir analizle ortaya koymuştur. Genel anlaşılırlık oranı %79,9 olup birçok öğrencinin ve kelimenin eşik değeri aştığı görülmüş; *yanlış kelime* tipi hatalar hem en yaygın hem de iletişimsel açıdan en riskli hata türü olarak belirlenmiştir. Literatürde zor kabul edilen bazı kelimeler anlaşılır çıkarken bazıları beklendiği gibi anlaşılmaz olmuştur; bu durum, "zor kelime" listelerinin bağlamdan bağımsız olarak genellenemeyeceğini göstermektedir. Bu bulgular, Dünya İngilizceleri ve ortak dil olarak İngilizce (ELF) perspektiflerini de desteklemekte olup birçok Türk aksanlı telaffuzun, temel fonemik karşıtlıklar korunduğu sürece anlaşılır kaldığını göstermektedir. Oturumlar arasında yaşanan ilerleme, tekrar eden üretimin olumlu etkisini vurgulamakta ve derslerde düşük riskli sözlü üretim görevlerinin teşvik edilmesini önermektedir. ASR teknolojileri, nesnel ölçüm imkânı sunarak özellikle kaynakların sınırlı olduğu bağlamlarda etkili bir tanılayıcı araç olabilir.

Telaffuz öğretimi, özellikle /θ/, /æ/, /iː/ gibi seslere odaklanacak biçimde daha yapılandırılmış hâle getirilmeli ve dil öğretmenlerinin ASR teknolojilerini tanıması ile kullanması desteklenmelidir. Sıkça yanlış telaffuz edilen kelimeler kurum bazında veriye dayalı olarak belirlenmeli ve öğrencinin kendi çıktısı üzerinden bireyselleştirilmiş geri bildirim verilmelidir. Bununla birlikte, ASR sistemlerinin insan dinleyicilerin fark ettiği algısal nüansları tam olarak yakalayamayabileceği ve bu çalışmanın yalnızca kelime düzeyiyle sınırlı kalıp vurgu, ritim ve tonlama gibi bürünsel özellikleri içermediği göz önünde bulundurulmalıdır. Gelecek çalışmalar, ASR çıktılarının insan değerlendirmeleriyle karşılaştırılmasını, tümce ve söylem düzeyindeki anlaşılabilirliğin incelenmesini ve farklı ASR motorlarının (örn. Google, Whisper, Azure) karşılaştırmalı analizini içermelidir.

Appendix A

Word	Freq.	Phonological Category	Word	Freq.	Phonological Category	Word	Freq.	Phonological Category
Multiple	7,6	Consonant clusters	*Slow*	7,4	Other	*Count*	10	Other
Probably	26,7	Consonant clusters	*Blow*	9,3	Other	*Duty*	6,2	Other
Procedure	5,4	Consonant clusters	*Separate*	6,2	Other	*Field*	17,1	Other
Present	12,7	Consonant clusters	*Below*	8,5	Other	*Busy*	7,5	Other
Capable	5,5	Consonant clusters	*Video*	11,8	Other	*Food*	17,3	Other
Comfortable	7,4	Consonant clusters	*Material*	11,2	Other	*Heart*	19,7	Other
Vehicle	6,1	Consonant clusters	*Event*	18,3	Other	*Beautiful*	12,3	Other
Prison	5,8	Consonant clusters	*Weapon*	7,4	Other	*During*	36,6	Other
Appropriate	7,1	Consonant clusters	*Listen*	17,7	Other	*Quarter*	7,1	Other
Photograph	6,5	Consonant clusters	*Connect*	8,8	Other	*Women*	29,7	Other
Worth	11,3	Interdental /θ/, /ð/	*Union*	7,7	Other	*Know*	64,2	Other
Think	57,6	Interdental /θ/, /ð/	*Conversation*	11,2	Other	*Air*	17,8	Other
Both	32,4	Interdental /θ/, /ð/	*Regular*	9,2	Other	*Kitchen*	6,8	Other
Thank	18,4	Interdental /θ/, /ð/	*Minute*	23,7	Other	*Session*	6	Other
Death	18,3	Interdental /θ/, /ð/	*Often*	28,5	Other	*Quote*	5,2	Other
Author	13,8	Interdental /θ/, /ð/	*Expect*	25,9	Other	*Develop*	16,7	Other
Breath	5,9	Interdental /θ/, /ð/	*Accept*	14,1	Other	*Question*	32,2	Other
Three	44,5	Interdental /θ/, /ð/	*Dead*	13,9	Other	*Suppose*	5,7	Other
Clothes	6,7	Interdental /θ/, /ð/	*Ear*	6,9	Other	*Colleague*	6,5	Other
Month	34,4	Interdental /θ/, /ð/	*Opportunity*	17,2	Other	*Foreign*	9	Other
Birth	6	Interdental /θ/, /ð/	*Could*	60,9	Other	*Answer*	15	Other
Those	55,4	Interdental /θ/, /ð/	*Area*	26	Other	*Suit*	6,1	Other
Health	15,8	Interdental /θ/, /ð/	*Coffee*	6,4	Other	*Wear*	16,2	Other
Thought	14,9	Interdental /θ/, /ð/	*Relation*	6,2	Other	*Determine*	13,2	Other
That	91,9	Interdental /θ/, /ð/	*Special*	19,1	Other	*Hotel*	6,8	Other
Although	21,4	Interdental /θ/, /ð/	*Success*	13	Other	*Via*	6,8	Other
These	56,3	Interdental /θ/, /ð/	*Decade*	15,1	Other	*Recognise*	13,3	Other
Thing	52,6	Interdental /θ/, /ð/	*Sign*	13,5	Other	*Two*	61,4	Other
Thousand	8,1	Interdental /θ/, /ð/	*Advice*	8	Other	*Because*	55,6	Other
Through	50,1	Interdental /θ/, /ð/	*Message*	13,5	Other	*Particular*	11	Other
Tooth	5,2	Interdental /θ/, /ð/	*Equal*	6	Other	*Tree*	9,7	Other
This	87,7	Interdental /θ/, /ð/	*Analysis*	8,4	Other	*Wednesday*	6,6	Other
Weather	6,2	Interdental /θ/, /ð/	*Ultimately*	7,5	Other	*Schedule*	5,2	Other
Breathe	5,1	Interdental /θ/, /ð/	*Equipment*	5,9	Other	*Ought*	5,2	Other
North	13,5	Interdental /θ/, /ð/	*Education*	12	Other	*Comment*	14	Other
The	99,5	Interdental /θ/, /ð/	*Communication*	7,4	Other	*Says*	73,4	Other
Mouth	8,8	Interdental /θ/, /ð/	*Wrong*	18,9	Other	*Would*	67,3	Other
Earth	10,2	Interdental /θ/, /ð/	*Effect*	17	Other	*Daughter*	11,8	Silent letters
Theory	7,8	Interdental /θ/, /ð/	*Head*	26,3	Other	*Doubt*	6,6	Silent letters
Live	34,1	Minimal pairs / multiple pronunciations	*Blood*	10,6	Other	*Write*	33,1	Silent letters
Lead	29,9	Minimal pairs / multiple pronunciations	*Low*	21,5	Other	*Island*	7,5	Silent letters
Wind	6,9	Minimal pairs / multiple pronunciations	*Occur*	12,4	Other	*Laugh*	9,8	Vowel complexity
Bear	7	Minimal pairs / multiple pronunciations	*Career*	12,9	Other	*Data*	11,4	Vowel complexity
Tear	5,4	Minimal pairs / multiple pronunciations	*Honor*	5,8	Other	*Affect*	11,7	Vowel complexity
Variety	8,6	Other	*Law*	19,2	Other			
Surface	7,2	Other	*Treat*	11,9	Other			

Appendix B

Chosen Turkish Target Words

Turkish Words				
Müteahhit	Laboratuvar	Müessese	Konsantrasyon	Poliklinik
Kağıt	Telaffuz	Muayenehane	Ansiklopedi	Enstitü

Appendix C

Individual Intelligibility Results of the Participants

Participant	S-1	S-2	S-3	TOTAL	%	Participant	S-1	S-2	S-3	TOTAL	%
G-1 P-01	86	94	96	276	**92**	**G-3 P-28**	68	69	87	224	**74,67**
G-1 P-02	66	72	67	205	**68,33**	**G-3 P-29**	66	73	74	213	**71**
G-1 P-03	75	78	71	224	**74,67**	**G-3 P-30**	79	84	87	250	**83,33**
G-1 P-04	83	86	90	259	**86,33**	**G-3 P-31**	91	87	96	274	**91,33**
G-1 P-05	80	86	90	256	**85,33**	**G-3 P-32**	74	78	78	230	**76,67**
G-1 P-06	66	80	75	221	**73,67**	**G-3 P-33**	63	71	71	205	**68,33**
G-1 P-07	64	66	83	213	**71**	**G-3 P-34**	68	71	68	207	**69**
G-1 P-08	66	78	69	213	**71**	**G-3 P-35**	52	62	65	179	**59,67**
G-1 P-09	86	80	80	246	**82**	**G-3 P-36**	70	88	90	248	**82,67**
G-1 P-10	73	77	83	233	**77,67**	**G-3 P-37**	63	76	79	218	**72,67**
G-1 P-11	72	78	77	227	**75,67**	**G-4 P-38**	97	91	99	287	**95,67**
G-1 P-12	91	94	90	275	**91,67**	**G-4 P-39**	70	69	66	205	**68,33**
G-1 P-13	91	87	93	271	**90,33**	**G-4 P-40**	80	78	80	238	**79,33**
G-2 P-14	96	97	90	283	**94,33**	**G-4 P-41**	65	69	75	209	**69,67**
G-2 P-15	87	90	92	269	**89,67**	**G-4 P-42**	76	77	86	239	**79,67**
G-2 P-16	79	86	91	256	**85,33**	**G-4 P-43**	81	84	87	252	**84**
G-2 P-17	84	92	82	258	**86**	**G-4 P-44**	82	77	81	240	**80**
G-2 P-18	73	82	85	240	**80**	**G-4 P-45**	78	93	95	266	**88,67**
G-2 P-19	63	70	66	199	**66,33**	**G-4 P-46**	71	79	82	232	**77,33**
G-2 P-20	81	89	91	261	**87**	**G-4 P-47**	79	79	95	253	**84,33**
G-2 P-21	79	73	79	231	**77**	**G-4 P-48**	97	89	91	277	**92,33**
G-2 P-22	81	79	84	244	**81,33**	**G-5 P-49**	84	76	76	236	**78,67**
G-2 P-23	83	82	79	244	**81,33**	**G-5 P-50**	80	82	86	248	**82,67**
G-2 P-24	77	86	88	251	**83,67**	**G-5 P-51**	84	83	87	254	**84,67**
G-2 P-25	74	75	82	231	**77**	**G-5 P-52**	71	81	79	231	**77**
G-2 P-26	80	81	88	249	**83**	**G-5 P-53**	65	73	70	208	**69,33**
G-2 P-27	88	83	82	253	**84,33**	**G-5 P-54**	81	76	72	229	**76,33**
TOTAL	2124	2221	2243			**TOTAL**	2035	2115	2202	12940	

* Each participant was evaluated on a total of 300 words across three sessions (100 words per session).
** G-1, G-2, and G-3 represent Group 1, Group 2, and Group 3, respectively.
*** Yellow indicates participants whose intelligibility exceeded the threshold when the error margin is included; orange indicates those who remained below the threshold in both raw and adjusted scores; green marks the participant with the highest overall intelligibility.