Beyond perceptual failure: cognitive load and number renditions in technology-mediated English–Turkish simultaneous interpreting

ENES EKİCİ and ŞEYDA ERASLAN

Dokuz Eylül University, Türkiye

This article investigates whether real-time simulated ASR visual support improves numerical rendition accuracy in English-Turkish simultaneous interpreting among student interpreters. Using a within-subjects experimental design, participants interpreted number-dense source texts with and without ASR-simulated captions. The findings are analyzed through Gile's Effort Model and Seeber's Cognitive Load Model. Results indicate that visual support did not eliminate numerical errors but shifted the error profile, reducing omissions while increasing syntactical errors – suggesting a redistribution of cognitive resources rather than an overall reduction in processing load.

Keywords: simultaneous interpreting, numerical rendition, computer-assisted interpreting, Cognitive Load Model, English–Turkish language pair, interpreter training

Introduction

Numbers have long been recognized as one of the most well-documented (Braun and Clarici 1996; Mazza 2001; Pinochi 2009; Mead 2015) sources of failure in simultaneous interpreting. Unlike lexical items, they resist prediction, offer little contextual redundancy and demand a kind of cognitive processing that sits uneasily alongside the already considerable demands of real-time interpreting. Gile's (1995) Effort Model frames this well: when Listening, Memory and Production efforts collectively exceed available capacity, something has to be sacrificed and numbers are frequently what gets sacrificed.

One response to this problem has been technological. The broader "technological turn" in interpreting (Fantinuoli 2018) has produced a generation of Computer-Assisted Interpreting (CAI) tools that move beyond preparation aids and glossary management to intervene directly during the interpreting task. The most recent of these tools use automatic speech recognition (ASR) to flag precisely the kind of high-density triggers, such as numbers, proper names, technical terms, that tend to push interpreters past their processing limits. Early experimental results are encouraging: studies have reported number rendition accuracy rising from under 60% to over 85% when real-time visual support is provided (Desmet et al. 2018; Defrancq and Fantinuoli 2021). However, most of this work has been conducted with professional interpreters or in language pairs other than English–Turkish, leaving open the question of how student interpreters respond to this kind of support.

The question this study sets out to answer is a practical one: does providing student interpreters with a visual display of numerical information actually help the rendition of numbers? And if so, how does it help—does the gain come simply from having the number visible, or does it reflect something deeper about how the brain handles multimodal input? Seeber's (2011) Cognitive Load Model offers a theoretical basis for optimism here. If visual-spatial and auditory-verbal channels operate as relatively independent resource pools, then a well-designed visual aid should relieve rather than compete with the auditory processing demands of interpreting. Whether that theoretical prediction holds in practice, specifically in the English–Turkish language pair and among student interpreters, is what this study examines.

English–Turkish interpreting presents a particularly interesting test case. The structural distance between the two languages means that even a correctly perceived figure can be incorrectly produced in the target language and this can happen in more than one way. Some figures map directly across the pair—English "4.2 trillion" corresponds to Turkish "4,2 trilyon", preserving both magnitude and structure—while others require structural reformulation: English reads years in paired two-digit units ("nineteen eighty-five"), whereas Turkish renders the full cardinal value ("*bin dokuz yüz seksen beş*"), demanding real-time decomposition rather than transcoding. A further difficulty concerns the order of magnitude rather than the digits themselves. English "million" and "billion" correspond to Turkish "*milyon*" and "*milyar*", near-identical forms that compete under time pressure, so that selecting the wrong scale word leaves the digits intact but relocates the entire figure across magnitudes. Although such a slip resembles a simple lexical substitution on the surface, the failure is structural—the number is reconstructed on the wrong scale. A figure can therefore be heard and understood correctly yet still fail at the production stage, for reasons that have nothing to do with perception. This distinction between perceptual failure and structural failure matters a great deal for how we understand what challenges in interpreting technology can and cannot address.

To investigate this, six senior interpreting students from the Department of Translation and Interpreting at a state university in Türkiye performed simultaneous interpretation of two numerically dense English speeches: one without any technological support and one accompanied by a simulated ASR display showing numbers in real time. The comparison allows for a direct assessment of whether visual support improves rendition accuracy and—equally importantly—whether it changes the type of errors students make.

By framing the current study within the broader context of the "technological turn" (Fantinuoli 2018), this research addresses the critical need for empirical data on how process-oriented tools affect the interpreter's cognitive performance in the English-Turkish language pair. While traditional strategies for managing numbers, such as the assistance of a boothmate or manual note-taking, have long been the standard, the integration of real-time visual support via Computer-Assisted Interpreting (CAI) represents a fundamental shift in the "cognitive resource footprint" of the task (Seeber 2017; Mazza 2001). Before detailing the experimental findings of the English–Turkish rendition, the following sections review the literature and theoretical frameworks that prop up this transition. This includes a historical analysis of how technology made simultaneous interpreting the new norm, the evolution of CAI generations and the cognitive models, moving from Gile's (1995) bottleneck theory to Seeber's (2011) multimodal approach.

Literature review and theoretical background

Technology and the Technological Turn in interpreting

Simultaneous interpreting as a professional practice is inseparable from the technology that made it possible. The booth, the headset, the relay system—these are not neutral tools but the conditions under which the profession came into existence. The Filene-Finlay system of the 1920s and its adaptation at the Nuremberg Trials (1945–1946) established that real-time multilingual communication was not only possible but scalable (Baigorri-Jalón 2014; Pöchhacker 2015). When the United Nations formally adopted simultaneous interpreting as a permanent service through Resolution 152 (II) in 1947, it did so for pragmatic reasons—time and efficiency—not because the profession had resolved its theoretical or cognitive questions (United Nations General Assembly 1947; Pöchhacker 2015).

What has changed in recent decades is the locus of technological intervention. Early technologies shaped the setting of interpreting: the booth, the transmission hardware, the acoustic environment. The current wave, which Fantinuoli (2018) terms the "technological turn", reaches further inward, targeting the cognitive sub-processes of the task itself. CAI tools now attempt to assist interpreters not before or after an assignment but during it, in real time. This is a qualitatively different kind of intervention, and it demands a correspondingly different theoretical framework for evaluation.

CAI tools and interpreter support during the task

The evolution of CAI tools is typically divided into three generations and the distinction between them is not merely technical but conceptual. First-generation tools—electronic glossaries like Interplex or Terminus—were essentially digitized reference materials, useful for preparation but passive during the task itself (Fantinuoli 2018). Second-generation tools introduced corpus-based preparation and smarter search functions, still oriented primarily toward the pre-task phase. The third generation is where the shift becomes genuinely significant: ASR-driven tools that operate in real time, surfacing high-density triggers, such as numbers, proper names and technical terms, at the moment they occur in the source speech (Fantinuoli 2018; Lu and Fantinuoli 2025).

The empirical case for these tools, at least where numbers are concerned, is difficult to dismiss. Desmet et al. (2018) reported number rendition accuracy rising from 56.5% to 86.5% with real-time visual support. Defrancq and Fantinuoli (2021) found similar gains with InterpretBank, with accuracy moving from 67.7% to 90.2% alongside a significant reduction in omissions. These are large effects by any standard and they suggest that the perceptual problem, failing to capture a number in the first place, is genuinely addressable through technology. What remains less clear and what this study attempts to prove, is whether solving the perceptual problem is sufficient, or whether structural and production-level failures persist even when the number is visible.

Numbers in simultaneous interpreting: Errors and strategies

The difficulty of numbers in simultaneous interpreting is well-documented and, at this point, largely uncontested. What is worth examining more carefully is why they are difficult, because the answer shapes what kind of support is actually useful.

Numbers are not simply unfamiliar words. They carry high information density with almost no contextual redundancy (Mazza 2001)—if you miss “forty-seven million”, there is no surrounding syntax that helps you reconstruct it. They also engage neural pathways that partially dissociate from those involved in ordinary language processing: exact numerical computation recruits both verbal memory networks and bilateral parietal areas associated with magnitude representation (Dehaene et al. 1999), a dual demand that has no real equivalent in lexical processing. Under Gile’s (1995) Effort Model, this cost is straightforward to describe: numbers intensify the Listening and

Short-term Memory efforts at precisely the moments when capacity is most constrained, pushing the system toward saturation.

Error rates reflect this. Braun and Clarici (1996) documented mean error rates approaching 70% among student interpreters. Professional interpreters fare better but remain far from reliable: Timarová (2012) reported error rates of approximately 40%, a figure consistent with the 43% observed at higher delivery rates by Desmet et al. (2018). The dominant error type in the unsupported conditions is omission—the interpreter drops the number rather than risk a more visible failure. In Pym's (2025) terms, this is a risk-management decision: the interpreter forgoes a high-risk item rather than produce a conspicuously erroneous figure. This is also consistent with Gile's framework: omission preserves the flow of the target speech at the cost of one piece of information. Approximation serves a similar function, rounding "47" to "around 50" to reduce local cognitive load while maintaining communicative continuity (Pinochi 2009).

The traditional human solution, a boothmate who notes figures as they appear, effectively externalizes the Memory Effort, offloading it to a visual aid. The importance of this collaboration was underscored by Arzık Erzurumlu and Demir (2022), who found that numerical accuracy declined sharply in Turkish broadcasts of the 2020 American Presidential Debates when COVID-19 social distancing requirements prevented interpreters from assisting each other in the booth. CAI tools are, in a sense, a technological formalization of this same principle—and one that does not depend on the physical proximity of a colleague.

Multimodality, ASR and cognitive load

The dominant theoretical framework for cognitive load in interpreting has long been Gile's Effort Model, and it remains indispensable as a descriptive tool. But it has a significant limitation: it treats attention as a single, undifferentiated resource. Under this view, any additional input—including a visual display—competes with the existing demands of listening and production and is therefore a potential source of overload rather than relief (Gile 2009).

This is where the model is demonstrably insufficient. The assumption of a unitary resource pool does not sit well with what we know about how the brain handles multimodal input. Wickens' (2002) Multiple Resource Theory proposes that the human information-processing system contains several relatively independent resource pools, organized by modality, processing code and stage. Two tasks interfere significantly only when they draw from the same pool; tasks that engage different modalities can, in principle, be performed concurrently at a much less cost.

Seeber (2011) builds directly on this foundation to offer a Cognitive Load Model specific to simultaneous interpreting. His "conflict matrix" makes the key point clearly: visual input is not inherently a competitor for the auditory-verbal resources that interpreting primarily demands. If a visual signal is synchronous and congruent with the auditory stream, as a live number display would be, it can function as cognitive support rather than additional load. This reframes the question entirely. The issue is not whether to introduce visual input, but how to design it so that it operates in a complementary rather than competing register.

Empirical support for this position is growing. Eye-tracking data suggest that interpreters strategically coordinate visual attention with auditory input during text-supported SI: professional interpreters show an ear-lead-eye pattern that enables efficient multimodal synchronization (Seeber et al. 2020), while trainee interpreters have been found to actively seek out numbers and proper names on screen when live captioning is available (Yuan and Wang 2023). ASR-generated captions have been shown to reduce cognitive load specifically for the kind of triggers, such as numbers and specialized terminology, where auditory processing is most vulnerable (Defrancq and Fantinuoli 2021). Taken together, this evidence supports Seeber's model over Gile's on this particular question: visual support, when well-designed, redistributes cognitive demand rather than compounding it.

Research questions

Guided by the theoretical frameworks of Gile (1995) and Seeber (2011) and the historical transition toward technology-mediated interpreting, this study addresses the following research questions:

RQ 1: What is the level of numerical rendition accuracy among student interpreters in the English–Turkish language pair?

RQ 2: Which strategies (e.g., omission, approximation, transcoding and delaying) do students employ when interpreting numbers?

RQ 3: Does the provision of technological support (simulated ASR) significantly affect the accuracy of numerical rendition?

Regarding RQ3, existing literature on multimodal facilitation suggests that technological support enhances performance by allowing for a "cognitive load transfer" between modalities. Seeber (2012) notes that interpreters actively seek visual stimuli to complement auditory speech, particularly when processing numerals. Experimental data from professional cohorts show that accuracy rises when supported by visual aids. Eye-tracking evidence further confirms that numerical processing imposes a disproportionately high cognitive load compared to surrounding content, as reflected in significantly longer fixation durations on numbers than on the items they refer to (Korpal and Stachowiak-Szymczak 2018). In one pilot study with trainees, number accuracy increased from 56.5% to 86.5% when supported by real-time visual displays (Desmet et al. 2018).

Methodology

This study employs a within-subjects experimental design to examine how real-time visual support affects the rendition of numerical information in English–Turkish simultaneous interpreting. Each participant served as their own control, interpreting one speech without technological assistance and one with simulated ASR support. This design was chosen to minimize the influence of individual variation in baseline interpreting ability, a particularly important consideration given the small sample size.

Participants

Six senior students (N = 6) from the Department of Translation and Interpreting at a major Turkish state university participated in the study. The experiment was conducted at the end of their eighth and final semester, ensuring that all participants had completed the full simultaneous interpreting curriculum, which includes Simultaneous Interpreting I and II in fall and spring terms respectively, each meeting 3 hours per week. All reported Turkish as their native language and English as their primary working language. The sample size is admittedly modest, and the findings should be read as exploratory evidence that warrants replication with a larger cohort rather than as definitive claims about the population of English–Turkish student interpreters.

Materials and stimuli

Two English-language speeches were prepared as experimental stimuli. The speeches were scripted to reproduce the register, structure and lexical density typical of conference presentations on the chosen topics, and were recorded by one of the researchers, a speaker with professional experience using English as a working language. Speech rate was controlled at 86 words per minute across both conditions. The target rate was achieved through the speaker’s rehearsed delivery during recording: no post-hoc audio manipulation was applied. The natural prosody, intonation and pause structure of the recording were therefore preserved. At 86 WPM, delivery falls below typical conference speech rates, which generally range from 100 to 120 WPM (Riccardi 2015). The reduced rate was a deliberate methodological choice. The study aimed to isolate the effect of ASR support on numerical rendition. Faster conference-typical rates (100-120 WPM) could have introduced a confounding variable, such as rate-induced cognitive overload. Holding speech rate constant at 86 WPM ensures that observed between-condition differences reflect the contribution of ASR support specifically. The ecological validity trade-off is acknowledged as a limitation and revisited in the Discussion.

Both speeches contained 59 numerical items. As with speech rate, this was a product of selection criteria rather than coincidence—speeches were chosen from a larger pool on the basis of comparable numerical density, topic neutrality relative to participant background knowledge and suitability for the target language pair.

Speech 1 (Control Condition): Topic: Global Warming. Duration: 9 minutes and 16 seconds. All the participants interpreted this speech without any technological support.

Speech 2 (Experimental Condition): Topic: E-commerce. Duration: 7 minutes and 47 seconds. All the participants interpreted this speech with simulated ASR support—a pre-recorded video display in which numerical information appeared on screen synchronized with the speaker's delivery. The display was not generated by a live ASR system; the numerical information was prepared in advance and assembled in iMovie as a pre-recorded video in which each figure appeared as white text on a black background, timed to appear on screen as the speaker articulated it. Prior to the experimental sessions, participants were provided with a terminology bank covering the lexical and topical content of both speeches, in line with standard interpreter training practice for advance preparation.

Numerical categorization

Before analysis could begin, all numerical items in both source texts were identified and classified. This step was necessary because not all numbers present the same cognitive challenge—a single-digit figure and a seven-digit figure make fundamentally different demands on the interpreter and collapsing them into a single category would obscure meaningful variation in the results.

The numbers appearing in the two source texts were grouped into five categories reflecting the types actually present in the speeches. This typology is informed by Jones's (2002) account of the processing dimensions that make numbers cognitively demanding—arithmetic value, order of magnitude, unit, extra-linguistic referent and relative value—on the premise that different number types draw on these dimensions to different degrees and therefore present distinct processing demands. Rather than adopting an existing scheme in full, the categories were selected to capture the specific items that occurred in this study. The resulting five-category scheme is as follows:

High-magnitude numbers refer to figures with four or more digits: thousands, millions, billions. These are consistently identified in prior literature among the most cognitively demanding numerical types (Mazza 2001; Pinochi 2009; Desmet et al. 2018).

Low-magnitude numbers cover single and double-digit whole numbers. These are generally less demanding but can still produce phonemic errors in English, particularly between near-homophones such as “thirteen” and "thirty." Dates include specific calendar years and historical references. Although structurally simple, dates are unforgiving—an error of even one digit changes the meaning entirely and there is no contextual redundancy to help the listener reconstruct the correct figure.

Percentages are numerical values expressed as a proportion of 100. They appear frequently in both speeches and carry high informational density relative to their length.

Decimals are non-percentage figures involving a decimal point. These proved to be the most resistant to improvement even with technological support, for reasons discussed in the results section.

Each numerical item in both source texts was assigned to one of these categories prior to analysis, producing a structured inventory against which the interpreted output could be systematically compared.

Error typology and data analysis

The interpreted recordings were transcribed in full and coded against the source text on a number-by-number basis. Each numerical item was first evaluated as either correct or incorrect. Items judged incorrect were then assigned to one of four error categories adapted from the typology developed by Pinochi (2009), which itself builds on the earlier framework of Braun and Clarici (1996). Following Pinochi, a lexical error preserves the order of magnitude while corrupting one or more digits (e.g., 277,000 → 276,000), whereas a syntactical error renders the wrong order of magnitude even when the figures themselves are correct (e.g., 47,000 → 47 million). This distinction is consistent with Frittella's (2019) coding of numerical errors. For the English–Turkish pair specifically, Bozok and Kıncal (2022) document numbers as a recurrent error source among student interpreters, situating omission within Pym's risk framework. Pinochi's original typology contains seven categories; three were excluded from the present study. The miscellaneous category was omitted on the grounds that unclassifiable errors cannot be meaningfully interpreted within either Gile's or Seeber's theoretical frameworks. Approximation and transposition, while well-documented in the literature, were not observed in either condition of this study and were therefore excluded from the analysis. The remaining four categories are retained as defined:

Omission occurs when a numerical item is entirely absent from the target output, or replaced by a non-specific expression such as "many" or "some." This is the most common error type in unsupported interpreting and functions, under Gile's (1995) framework, as a tactical sacrifice—the interpreter drops the number to preserve the coherence of the surrounding speech. Although it is generally considered a tactical sacrifice, according to Pym (2008), interpreters also make omissions based on their evaluation of the communicative risk. Thus, low risk omissions "are part of a general economy of time management, mostly as part of a general strategy of implicitation" (Pym 2008: 95). Therefore, "interpreters might make omissions as a time-saving strategy for pragmatic reasons unless the omitted information is crucial for the purposes of the communicative act", which requires considering the contextual dimension of omissions besides the cognitive (Kıncal 2020: 99).

The concept of risk management in translation (including interpreting) by Anthony Pym is highly relevant in this context (Pym 2025). With a broader view of interpreting as a form of intercultural communication, the strategies that orient the decisions of interpreters from cognitive to social are defined by their intuitive assessments of the "effects that each decision will have on the imagined receiver", which can be considered as "embodiments of successful risk management decisions" (2025: 3). Thus, numbers can be sacrificed as a risk management strategy rather than a high-risk error.

Lexical error occurs when the order of magnitude is correct but specific digits or components are changed—"346" rendered as "436", for example. These errors suggest that the interpreter captured the scale of the number but lost precision at the digit level.

Syntactical error occurs when the digits are correct, but the order of magnitude is rendered incorrectly—"47" produced as "470", or "47 million" instead of "47 thousand." In the English–Turkish context, this error type is particularly significant because the two languages differ in how they structure large numbers, creating a production-level challenge that persists even when the source figure has been correctly perceived.

Error of phonemic perception results from the phonological similarity of certain English figures—most commonly the -teen/-ty distinction, as in "fourteen" versus "forty." These errors are a source feature of the English input rather than a target language production failure.

Results

This section presents the findings of the comparative analysis between the unsupported condition (Speech 1) and the ASR-supported condition (Speech 2). Results are organized across three dimensions: global accuracy and individual performance, accuracy gains by numerical category and the shift in error profiles between conditions.

Accuracy and individual performance

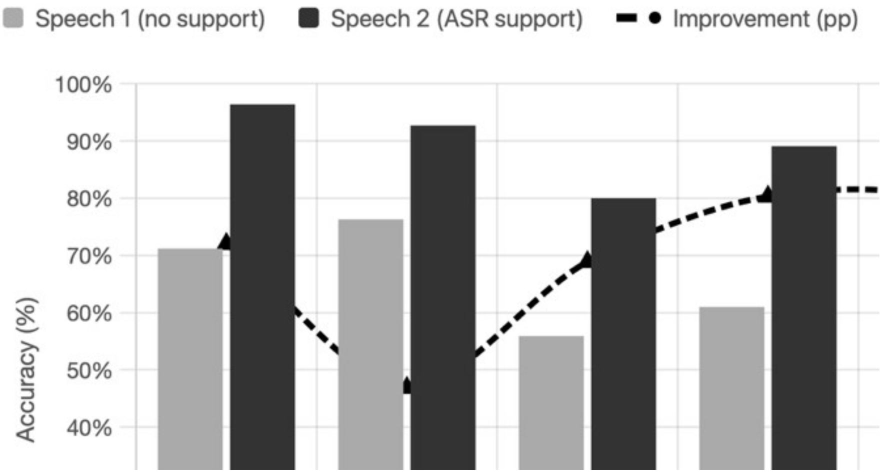
Across all six participants, accuracy improved in the ASR-supported condition without exception. In Speech 1, individual accuracy rates ranged from 54.2% (INT5) to 76.3% (INT2), with a cohort mean of 64.4%. In Speech 2, the range shifted upward from 80.0% (INT3) to 96.4% (INT1), with a cohort mean of 87.6%. The magnitude of individual improvement varied considerably. INT2, who entered the experimental condition with the highest baseline, showed the smallest gain at 16.5 percentage points. INT4, by contrast, improved by 28.1 points—the largest individual gain in the cohort. INT1, INT3 and INT5 showed improvements of 25.2, 24.1 and 27.6 points respectively, clustering around a similar range. INT6 improved by 17.7 points, placing alongside INT2 at the lower end of the improvement distribution.

The individual figures are as follows:

Table 1: Accuracy and individual performance

	Speech Accuracy 1	Speech Accuracy 2	Improvement
INT1	71.2%	96.4%	+25.2 pts
INT2	76.3%	92.7%	+16.4 pts
INT3	55.9%	80%	+24.1 pts
INT4	61%	89.1%	+28.1 pts
INT5	54.2%	81.8%	+27.6 pts
INT6	67.8%	85.5%	+17.7 pts

Figure 1: Individual accuracy rates (%) in Speech 1 and Speech 2, with improvement in percentage points (right axis).



Two observations are worth noting at this stage. First, the two participants with the smallest gains—INT2 and INT6—were also among those with the higher baselines, suggesting a ceiling effect may have constrained their improvement. Second, the participants with the lowest baselines showed the steepest gains, indicating that the visual support was especially consequential for those who struggled most in the unsupported condition.

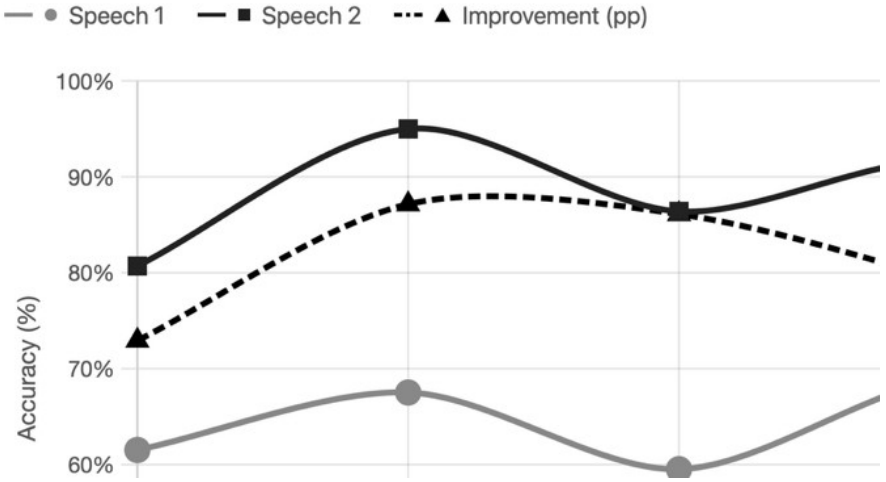
Accuracy gains by numerical category

The improvement was not uniform across numerical types. Table 2 presents the cohort-level accuracy figures for each category across both conditions.

Table 2: Accuracy gains by numerical category

Number Type	Speech Accuracy 1	Speech Accuracy 2	Improvement
High Magnitude	61.5%	80.7%	+19.2 pts
Percentages	67.5%	95%	+27.5 pts
Dates	59.5%	86.4%	+26.9 pts
Low Magnitude	69%	91.7%	+22.7 pts
Decimals	66.7%	83.3%	+16.6 pts

Figure 2: Cohort-level accuracy (%) by numerical category across both conditions, with improvement in percentage points (right axis).



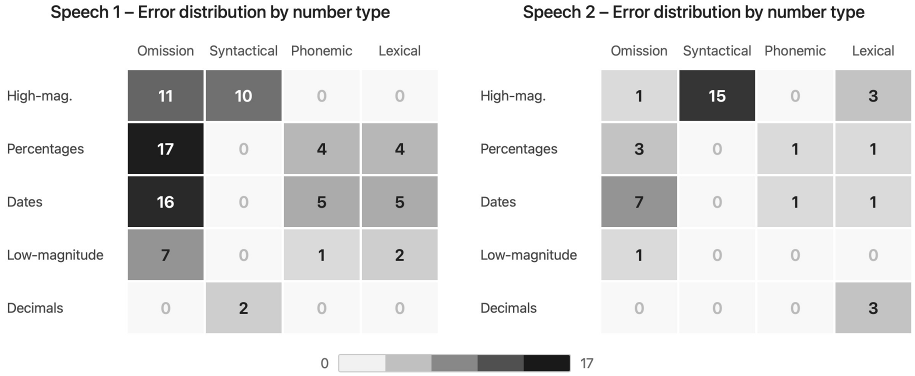
Percentages showed the largest absolute gain at 27.5 percentage points, reaching a near-ceiling accuracy of 95.0% in the supported condition. Dates followed closely with a 26.8-point improvement. Both categories share a structural feature that likely explains their responsiveness to visual support: they are high in informational density but relatively standard in format, meaning that once the figure is visible, its integration into the target output is comparatively straightforward.

High-magnitude numbers showed a more modest improvement of 19.2 points, remaining the lowest-accuracy category in both conditions. This is significant: despite the visual display making the figure available, participants still struggled with these items more than any other types. Decimals showed the smallest gain at 16.7 points, suggesting that even with visual support, this category presents particular challenges—a point returned to in the discussion.

Error profile shift

The heatmap data reveal not just a reduction in errors but a fundamental change in the type of errors produced, which is arguably the more theoretically significant finding. Notably, approximation—a strategy well-documented in the literature—was effectively absent in both conditions, suggesting that the participants in this study either committed to a full rendition or dropped the number entirely.

Figure 3: Error distribution by number type and error category for Speech 1 (left) and Speech 2 (right).



Note: Cell values indicate total error count across all six participants. Darker shading reflects higher frequency.

In Speech 1, omission was the dominant error type across virtually all categories. Percentages produced 17 omissions, Dates 16 and High-magnitude numbers 11—together accounting for the large majority of all errors in the unsupported condition. Syntactical errors were present but secondary, concentrated in High-magnitude numbers (10 instances). Phonemic errors appeared in moderate numbers across Percentages, Dates and High-magnitude items. Lexical errors were confined to Percentages, Dates and Low-magnitude numbers, with 4, 5 and 2 instances respectively.

In Speech 2, this profile changed substantially. Omissions decreased substantially across most categories: Percentages fell from 17 to 3, Dates from 16 to 7 and High-magnitude numbers from 11 to 1. Phonemic errors similarly declined. However, syntactical errors in the High-magnitude category increased sharply, rising from 10 to 15—the only error type to worsen in the supported condition. Lexical errors in the High-magnitude category also rose slightly, from 0 to 3.

The pattern is clear: ASR support largely solved the perceptual problem. Participants who previously dropped numbers entirely were now capturing them from the visual display. What the support did not resolve—and in some respects made more visible—was the production problem. With the figure now in hand, participants still had to construct it in Turkish, and it is at this structural integration stage that errors concentrated in the supported condition.

Discussion

The results of this study invite a reconsideration of how cognitive load is theorized in simultaneous interpreting—not a wholesale rejection of existing frameworks, but a refinement of what they can and cannot explain when technological support is introduced into the interpreting task at hand.

Gile's Effort Model remains a productive starting point. The baseline data from Speech 1 maps cleanly onto its central prediction: when the combined demands of Listening, Memory and Production exceed available capacity, the system sheds load by the path of least resistance. In numerically dense speech, that path is omission. The high omission rates observed across all participants in unsupported condition, 17 instances for percentages alone, 16 for dates, are precisely what Gile's framework would lead us to expect. Interpreters were not making careless errors; they were making triage decisions under pressure, sacrificing individual figures to preserve the coherence of the surrounding speech—what Pym (2025) characterizes as the management of communicative risk.

Where the model shows its limits is in accounting for what happened next. Under a strictly unitary resource view, introducing a visual display during interpreting should add to the cognitive burden, another input stream competing for the same finite pool of attention. The data suggests otherwise. Accuracy improved universally across all six participants and omissions collapsed in every category where the visual support was available. These results suggest the visual support relieved cognitive load rather than adding to it. This is where Seeber's (2011) Cognitive Load Model offers a more adequate account. If visual-spatial and auditory-verbal processing draw on partially independent resource pools, as Wickens' (2002) Multiple Resource Theory proposes, then a congruent visual display does not compete with the auditory stream—it complements it. The number on screen reaches the interpreter through a different channel than the number in the audio and the two together are less costly than either alone under saturating conditions.

That said, the data also complicates any straightforward celebration of multimodal support, and this is where the study's most significant finding lies.

The shift in error profile between Speech 1 and Speech 2 is not merely a quantitative improvement—it is a qualitative transformation in the nature of failure. In unsupported condition, the dominant error was omission: the interpreter did not have the number and could not produce it. In the supported condition, omission was largely eliminated, but syntactical errors in the high-magnitude category increased from 10 to 15 instances. Interpreters now had the number—they could see it—but they still struggled to render it correctly in Turkish. The problem had moved downstream, from perception to production.

This distinction matters theoretically. Gile's model treats cognitive saturation as a single phenomenon; the data suggests it has at least two distinct stages. The first is perceptual saturation, the failure to capture a stimulus under auditory overload. Technology, in this study, effectively solved that problem. The second is structural saturation, the failure to integrate a correctly perceived stimulus into the morphosyntactic architecture of the target language under time pressure. Technology did not solve that problem and arguably made it more visible by removing the perceptual problem that had previously masked it.

The structural challenge is particularly acute in English–Turkish interpreting, where production failures can arise independently of perception. The clearest case in the present data involves the order of magnitude: English "million" and "billion" correspond to Turkish "*milyon*" and "*milyar*", near-identical forms that compete under time pressure. When an interpreter sees "2.3 billion" on screen, the perceptual task is complete—the figure has been correctly received—yet selecting the right scale word in production remains demanding, and an error here relocates the entire figure across magnitudes while leaving the digits intact. The persistence of syntactical errors in the supported condition suggests that this production-level bottleneck is not addressable through visual input alone: seeing the number does not resolve the competition between scale words that arises at the moment of production.

The category-level results add further nuance to the multimodal facilitation picture. Percentages and dates showed the largest gains—27.5 and 26.8 percentage points respectively—while decimals showed the smallest at 16.7 points. This variation is not random. Percentages and dates are structurally simple in Turkish: once the figure is captured, its integration into the target output is relatively formulaic. Decimals, by contrast, involve notational complexity that requires cognitive investment at the production stage regardless of how the figure was perceived. The visual display helps most where the production task is straightforward; it helps least where production itself is the bottleneck. This is consistent with Seeber's (2017) observation that visual input functions as a cognitive aid specifically when it offloads a process that would otherwise compete with production—not as a universal solution to all sources of difficulty.

Taken together, these findings suggest a more granular picture of the "augmented interpreter" than the literature has so far offered (Fantinuoli and Dastyar 2022). Technology does not simply reduce cognitive load—it redistributes it. The interpreters who work with ASR support are no longer primarily managing a perceptual problem; they are managing a structural one. This is a meaningful shift, but it is not a resolution. The booth becomes a different cognitive environment, not an easier one.

Conclusion

The findings of this study converge on a point that is both empirically grounded and theoretically generative: technology shifts the interpreter's failure mode rather than eliminating it. This is not a pessimistic conclusion, the accuracy gains were real and consistent, but it is a clarifying one.

The baseline condition confirmed what the literature on numerical processing in SI has long established. Numbers are among the most frequent sources of failure in interpreting performed without in-booth technological assistance and omission is the dominant response - in some cases, the strategy - to that failure. Student interpreters working without technological assistance sacrificed figures at high rates, particularly percentages, dates and high-magnitude numbers, not out of incompetence but out of rational prioritization under pressure. Gile's Effort Model describes this dynamic well and the Speech 1 data offer straightforward empirical support for its core prediction.

The introduction of simulated ASR support changed the picture substantially. Accuracy improved for every participant without exception and omissions declined sharply across nearly all numerical categories. The visual display allowed interpreters to bypass the auditory bottleneck that had previously made numerical capture so costly. This is the outcome Seeber's multimodal framework predicts and the consistency of the effect across participants and categories gives it credibility even within the constraints of a small exploratory sample.

But the more significant finding is what happened to the nature of failure. With the perceptual problem largely resolved by visual support, structural errors became the dominant failure mode. Interpreters who could now see the figure still struggled to render high-magnitude numbers correctly in Turkish, where the syntactic demands of production remain considerable regardless of how the stimulus was received. Technology solved one bottleneck and exposed another.

This points toward something the field has perhaps underappreciated in its enthusiasm for AI-assisted interpreting. Aggregated accuracy metrics, the kind routinely reported in CAI studies, can tell us what the results of technological intervention are, but they are less equipped to tell us how those results come about, or at what cognitive cost. The distinction between Gile and Seeber is not merely a theoretical disagreement; it is a methodological one. A unitary resource model and a multimodal model make different predictions about where failure will concentrate, and those predictions have practical consequences for how we design tools, evaluate performance and train interpreters. Without that theoretical granularity, we risk building an “augmented interpreter” model on foundations that do not fully account for the cognitive architecture of the task.

The English–Turkish language pair has received comparatively little attention in prior literature (apart from Arzık Erzurumlu and Demir 2022) and this study is exploratory in scope. But it suggests that the pair presents specific challenges, particularly in the structural integration of high-magnitude numbers, that deserve dedicated investigation. More broadly, it adds to the growing body of evidence that simultaneous interpreting cannot be adequately understood through auditory-verbal models alone. The booth has always been a multimodal environment; what AI-backed technologies have done is make that fact impossible to ignore.

AI acts as a facilitator; yet the degree to which it does so and the cognitive costs it redistributes rather than removes, remain open questions that warrant further investigation.

Declaration of AI use

In preparing this manuscript, the authors used Claude for preparing the tables. All AI-assisted content was reviewed, edited and verified by the authors, who take full responsibility for the accuracy and integrity of the final text.

References

Arzık Erzurumlu, Özüm, and Demir, Perihan. 2022. “Taking stock of the COVID-19 working conditions on the performance of the interpreters: Rendering the numbers in the 2020 American Presidential Debates”. *Journal of Translation Studies* 32: 25–54.

Baigorri-Jalón, Jesús. 2014. *From Paris to Nuremberg: The Birth of Conference Interpreting.* Translated and edited by Holly Mikkelson and Barry Slaughter Olsen. Amsterdam and Philadelphia: John Benjamins.

Bozok, Nazlıgül, and Kıncal, Şeyda. 2022. “Strategies and errors in simultaneous interpreting: A student-oriented experiment in English–Turkish language pair”. *Current Trends in Translation Teaching and Learning E*, 32–75. https://doi.org/10.51287/cttl20222

Braun, Sabine and Clarici, Andrea. 1996. “Inaccuracy for numerals in simultaneous interpretation: Neurolinguistic and neuropsychological perspectives”. *The Interpreters' Newsletter* 7: 85–102.

Defrancq, Bart, and Fantinuoli, Claudio. 2021. “Automatic speech recognition in the booth: Assessment of system performance, interpreters' performances and interactions in the context of numbers”. *Target* 33(1)*:* 73–102.

Dehaene, Stanislas, Spelke, Elizabeth S., Pinel, Philippe, Stanescu, Ruxandra, and Tsivkin, Sanna. 1999. "Sources of mathematical thinking: Behavioral and brain-imaging evidence". *Science,* 284(5416): 970–974. https://doi.org/10.1126/science.284.5416.970

Desmet, Bart, Vandierendonck, Mieke, and Defrancq, Bart. 2018. "Simultaneous interpretation of numbers and the impact of technological support". In C. Fantinuoli (Ed.), *Interpreting and Technology.* Berlin: Language Science Press. 13-27.

Fantinuoli, Claudio. 2018. "Interpreting and technology: The upcoming technological turn". In C. Fantinuoli (Ed.), *Interpreting and Technology*. Berlin: Language Science Press. 1-12.

Fantinuoli, Claudio and Dastyar, Vorya. 2022. "Interpreting and the emerging augmented paradigm". *Interpreting and Society: An Interdisciplinary Journal* 2(2): 185–194.

Frittella, Francesca Maria. 2019. "70.6 billion world citizens': Investigating the difficulty of interpreting numbers". *Translation & Interpreting* 11 (1), 79–99. https://doi.org/10.12807/ti.111201.2019.a05

Gile, Daniel. 1995. *Basic Concepts and Models for Interpreter and Translator Training.* Amsterdam and Philadelphia: John Benjamins.

Gile, Daniel. 2009. *Basic Concepts and Models for Interpreter and Translator Training* (Rev. ed.). Amsterdam and Philadelphia: John Benjamins.

Jones, Roderick. 2002. *Conference Interpreting Explained.* Manchester: St. Jerome Publishing.

Kıncal, Şeyda. 2020. "The impact of non-native English on omissions in simultaneous interpreting". *Journal of Translation Studies* 28: 98-109.

Korpal, Paweł, and Stachowiak-Szymczak, Katarzyna. 2018. "The whole picture: Processing of numbers and their context in simultaneous interpreting". *Poznan Studies in Contemporary Linguistics* 54(3): 335–354.

Lu, Xinchao, and Fantinuoli, Claudio. 2025. "Machine and computer-assisted interpreting: Innovations in and implications for interpreting practice, pedagogy and research". *Linguistica Antverpiensia, New Series – Themes in Translation Studies,* 24. https://doi.org/10.52034/lans-tts.v24i.869

Mazza, Cristina. 2001. "Numbers in simultaneous interpretation". *The Interpreters' Newsletter* 11: 87–104.

Mead, Peter. 2015. "Numbers". In Franz Pöchhacker (ed.), *Routledge Encyclopedia of Interpreting Studies*, 286–288. Abingdon and New York: Routledge.

Pinochi, D. 2009. Simultaneous Interpretation of Numbers: Comparing German and English to Italian. An Experimental Study. *The Interpreters' Newsletter*, *14*, 33–57. http://hdl.handle.net/10077/3465

Pöchhacker, Franz. 2015. *Introducing Interpreting Studies* (2nd ed.). New York and London: Routledge.

Pym, Anthony. 2008. “On omission in simultaneous interpreting: Risk analysis of a hidden effort”. In G. Hansen et al. (eds) *Efforts and Models in Interpreting and Translation Research: A Tribute to Daniel Gile.* Amsterdam and Philadelphia: John Benjamins. 83-105.

Pym, Anthony. 2025. *Risk Management in Translation*. Cambridge: Cambridge University Press.

Riccardi, Alessandra. 2015. “Speech rate”. In F. Pöchhacker (Ed.), *Routledge Encyclopedia of Interpreting Studies*. London: Routledge. 397–399.

Seeber, Kilian G. 2011. “Cognitive load in simultaneous interpreting: Existing theories—New models”. *Interpreting* 13(2): 176–204.

Seeber, Kilian G. 2012. “Multimodal input in simultaneous interpreting: An eye-tracking experiment”. In L. N. Zybatow, A. Petrova, and M. Ustaszewski (Eds.), *Proceedings of the 1st International Conference TRANSLATA*. Frankfurt am Main: Peter Lang. 341–347.

Seeber, Kilian G. 2017. “Multimodal processing in simultaneous interpreting: A multiresource model”. In J. W. Schwieter, A. Ferreira (eds), *The Handbook of Translation and Cognition*. Hoboken, New York: Wiley. 1-16.

Seeber, Kilian G., Keller, Laura, and Hervais-Adelman, Alexis. 2020. “When the ear leads the eye – the use of text during simultaneous interpretation”. *Language, Cognition and Neuroscience* 35(10): 1480–1494.

Timarová, Šárka. 2012. *Working Memory in Conference Simultaneous Interpreting.* Unpublished doctoral dissertation. Charles University/KU Leuven.

United Nations General Assembly. 1947. Resolution 152 (II): Simultaneous Interpretation. A/RES/152(II), 15 November 1947.

Wickens, Christopher D. 2002. “Multiple resources and performance prediction”. *Theoretical Issues in Ergonomics Science* 3(2)*,* 159–177.

Yuan, Lu, and Wang, Binhua. 2023. “Cognitive processing of the extra visual layer of live captioning in simultaneous interpreting: Triangulation of eye-tracked process and performance data”. *Ampersand* 11, 100131.