It’s the Timing, Not the Script: How Delivery Drives Engagement

By Alex Tan · Published 18 November 2026



What makes a script sound like it's being read? It's mostly timing: how fast the words come and how evenly the pauses fall. And how it's delivered. In controlled studies, the same words delivered differently changed attention, trust and persuasion.

Download this paper as a PDF (the same text, with the diagrams and references).

What the best-remembered speeches have in common

In 1999, 137 scholars of American public speaking ranked the century's best speeches, judged on their artistry and on what they changed. First place went to Martin Luther King Jr.'s "I Have a Dream", given at the March on Washington in 1963 [1]. CBS carried that whole afternoon live [2].

The words people remember weren't in the text he was reading. "I started out reading the speech," King said later. Then "all of a sudden this thing came to me", and "I just turned aside from the manuscript altogether" [3]. The dream passage, the part every clip plays, came after he stopped reading.

Three years earlier, Kennedy and Nixon met in the first of the 1960 televised debates. The story goes that people who watched thought Kennedy won, and people who listened on the radio thought Nixon did. That was hard to check until a 2003 experiment replayed the debate to 171 people who didn't know it.

Of those 171, 85 watched and 86 heard only the sound. 81% of the viewers judged Kennedy the winner, against 60% of the listeners [4]. Viewers also leaned more on how honest each man seemed. Same words, same voices. What changed was what people could see. (They were mostly young university students, in one lab study.)

Neither example says scripts are bad. King had a manuscript, and both candidates had prepared lines. What set the moments apart was delivery: King speaking instead of reading, and what viewers saw on top of what they heard.

That's the subject of this paper. It's for anyone who reads to camera from a script. Much of what makes a read sound read happens between the words: the pace, and where the pauses fall.

What ”engagement” means here

"Engagement" gets used for everything from likes to comments to watch time. In this paper it means four outcomes that studies of delivery have actually measured:

- **Attention.** Whether listeners stay with the message and remember its words, measured with recall, recognition and arousal [5].
- **Trust.** Whether listeners believe what the speaker says [6].
- **Persuasion.** Whether listeners' attitudes and choices move towards the speaker's [7].
- **Watch time.** How long viewers keep watching. There's one large dataset, from online course video [8].

Each comes from a different kind of study. None of the studies looked at creator videos on YouTube or TikTok. They used voices, faces and lecture videos.

However, there are two things that are unknown. We don't know if better delivery helps viewers understand or learn more [9] [10]. And we don't know if better delivery helps viewers retain the information.

What listeners hear when a script sounds read

Most people can tell. In one study, Swiss listeners heard short clips of people either reading aloud or speaking freely in their local dialect, and picked which was which well above chance [11]. In another, Dutch speakers gave instructions off the cuff, then read their own words back from a transcript. Listeners told the two apart 77 to 79% of the time [12].

But what are they hearing? It isn't mainly pitch. The usual advice to "vary your tone" assumes read speech is flat and free speech is lively, and the data don't line up that neatly. In the Dutch study, the pitch differences between reading and speaking went in opposite directions in two different sets of recordings [12].

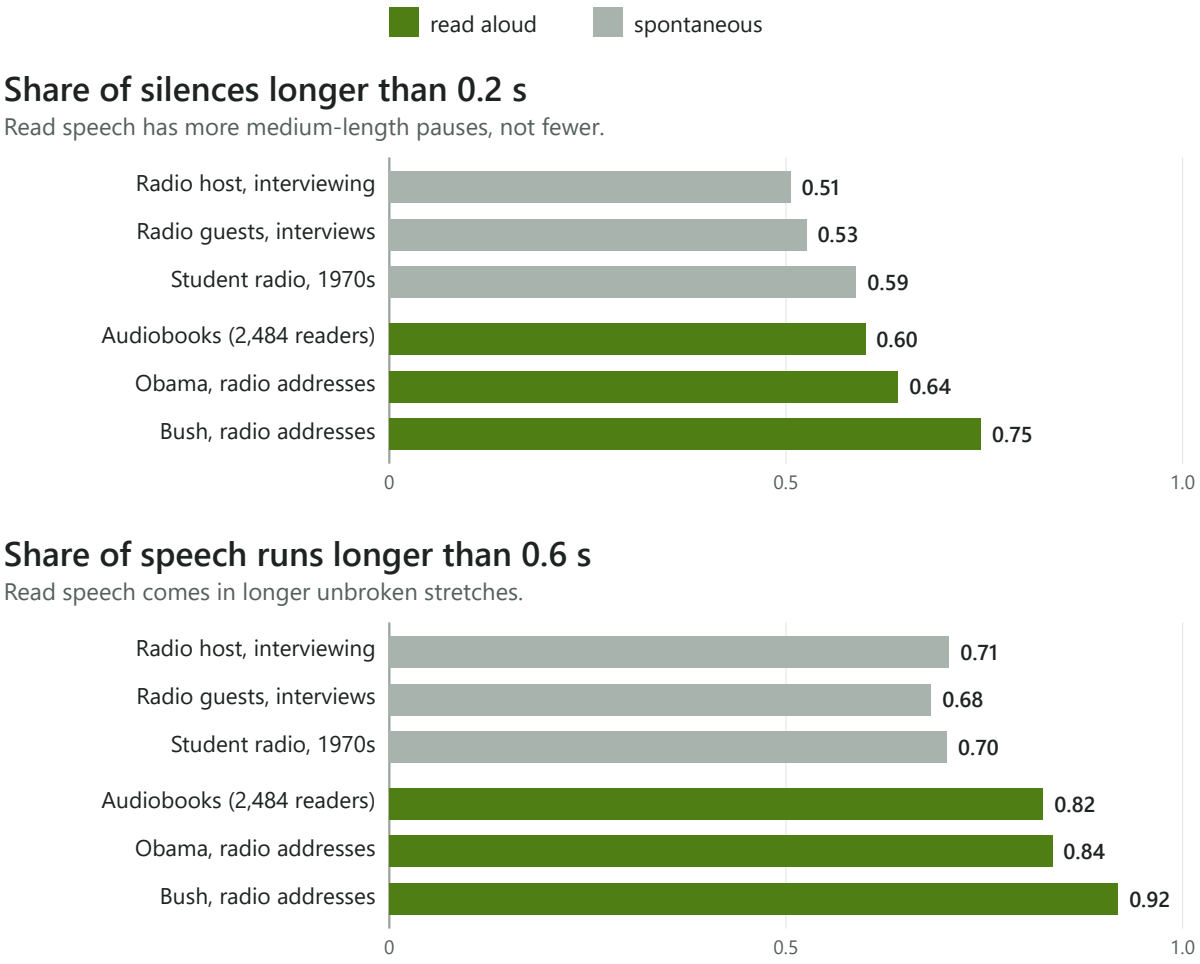
In a large English study, pitch range varied widely within each style. Two speakers reading prepared addresses used ranges of 11.5 and 6.8 semitones, and two interview guests 10.3 and 6.8 [13].

Timing is the steadier cue. In the Swiss study, eight prosodic measures were tested against what listeners decided. Only one explained their judgements: articulation rate, the speed of the speech between the pauses [11]. In the English study, the clearest difference between the styles was in how long the stretches of speech and silence were [13].

What that difference looks like surprises people. Read speech doesn't pause less. It pauses more evenly.

In presidents' weekly radio addresses and in audiobooks, 60 to 75% of silences lasted longer than a fifth of a second, against 51 to 59% in radio interviews. The read speech also came in longer unbroken runs [13].

Free speech spreads out both ways, with more very short gaps and a tail of long ones. In a French comparison, pauses longer than a second turned up only in spontaneous speech. Pauses over two seconds made up 2.8% of the spontaneous ones and none of the read [14].



Data: Ryant & Liberman (2016), Table 2. American English. Different speakers in each style, so read the pattern, not the gaps.
In French recordings, pauses over 2 s were 2.8% of spontaneous pauses and 0% of read ones (Campione & Véronis, 2002).

So a read voice has a recognisable rhythm: steady runs, with middling pauses at regular places. Someone thinking aloud has a ragged one, with a quick catch here and a long beat there, wherever the thought needs it.

These studies have limits. They cover Swiss German, Dutch, French and American English, with short clips and lab tasks, and the corpus studies compare different speakers. The authors of the English study also reject a strict line between read and spontaneous speech [13].

And with identical words, the gap narrows. In an English study where the read and spontaneous versions were word for word the same, listeners who came in cold were at chance. Those who had done the speaking task themselves first did better [15]. So the difference can be subtle, and the ear for it comes partly from practice.

Same words, different delivery

If timing is what makes a script sound read, does sounding read cost you anything? No study has tested that chain end to end. But several have held the words still and changed the delivery, and it made a difference.

Persuasion. In five experiments, speakers read the same product review twice, once normally and once trying to persuade, without changing a word [7]. Over 4,000 online listeners heard the recordings. The persuasive readings moved attitudes, and in a choice task raised the share picking the product from 76% to 81%. The effect ran through how confident the speaker sounded.

The effects were small, and the cues that carried them were volume and how much it varied. Speaking rate and pauses made no difference to persuasion in that study. So it shows delivery persuades. It doesn't show timing does.

Attention. Radio ads recorded with a varied pitch range, rather than a flat one, drew more attention and arousal, and listeners recalled and recognised more of the words [5]. That's audio only, and it's about pitch, not timing. So pitch still matters for holding attention, even though it isn't what gives a read away.

Watch time. Across 6.9 million viewing sessions of online course videos, faster speakers tended to hold viewers longer, up to twice as long for videos of a similar length [8]. The authors took speed as a sign of enthusiasm, which they judged by watching a sample of videos rather than measuring it.

It's course video from 2012, not creator content, and the line isn't clean. For 6 to 12-minute videos, mid-speed speakers did worse than the slowest. The authors' advice was to reassure instructors that speaking fast is okay.

Trust. Delivery is also where your eyes go. Listeners believed truth-ambiguous statements more often when the speaker looked into the camera: 6.71 of 12, against 5.2 to 5.5 when she looked away [6]. That was one speaker and 35 listeners. With the sound alone, the difference disappeared.

Across 18 studies of instructional video, an instructor's direct gaze improved learning a little ($g = 0.30$) and raised viewers' sense of a personal connection ($g = 0.34$) [16]. It also drew their eyes away from the material on screen.

Reading, heard as reading. One study came close to this paper's subject. Over a 31-minute lecture, the same instructor either taught fluently or read from notes in a monotone, with awkward pauses. Students rated the reader as less effective [10].

Two findings cut the other way, though:

- **Fluency feels like learning.** In that lecture study and an earlier one-minute version, students who watched the fluent instructor believed they'd learned more. Yet, their memory tests didn't differ [9] [10]. Delivery changes how a video is judged. On this evidence, it doesn't change how much is understood.
- **Eye contact can backfire.** Viewers who already disagreed with a speaker were persuaded less when the speaker looked at them [17].

Pacing to the machine

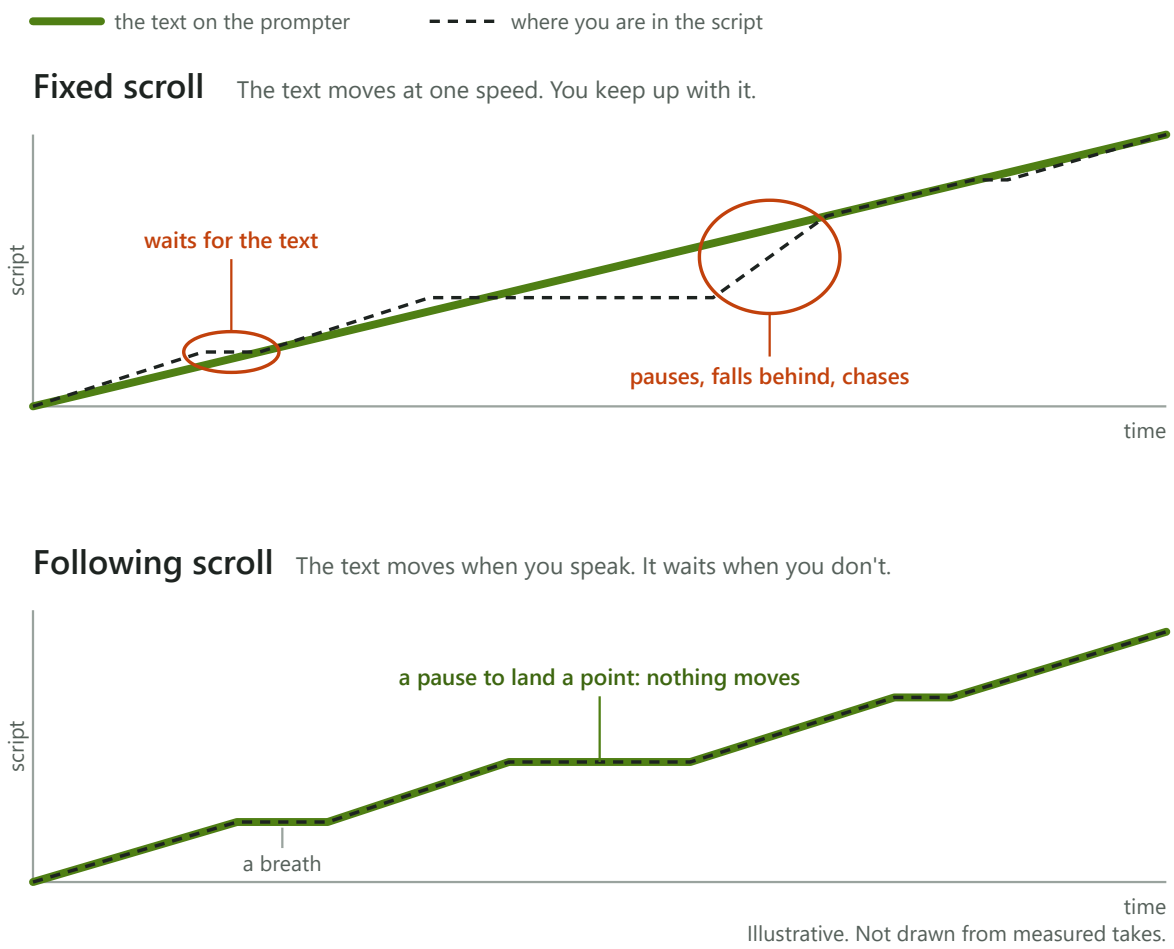
A fixed-speed prompter moves the text at a constant rate. But your speech never does that. It comes in bursts, a phrase at a time, with silences between them that change with what you're saying. A short beat after a list item. A longer one before the point you want to land. A breath where the sentence runs out.

A constant scroll can't match bursts. So at every phrase you do one of two things:

1. If the text is behind you, you wait for it, and the pause falls where the scroll brings up the next words, not where the sentence needs one.
2. If the text is ahead, you chase it. You cut the pause short or skip it, and rush the next phrase to get back under the reading line.

Either way, the timing in the take is now in the hands of the scroll. That means, you are **pacing to the machine**: your rate and your pauses follow the prompter's clock, instead of the meaning of your words.

It would explain something odd about prompter reads. The script can be written for the ear and delivered with plenty of energy, and still sound read. If listeners hear "read" mostly in timing, as the research above suggests, then pacing to the machine produces the very cue they listen for, whatever else you get right.



Someone used to do the cue cards

Prompting didn't start at a fixed speed. Cue cards came to television in 1949, when Barney McNulty wrote out Ed Wynn's lines on large cards [18]. The first teleprompter, used on CBS's *The First Hundred Years* in 1950, was a paper scroll in a box by the camera, run from off stage by a trained technician [19].

A 1953 photo of one in use has the caption: "Operator, at right, controls speed of device to correspond with star's delivery" [20].

When the operator didn't follow, people noticed. At the 1952 Republican convention, the operator slowed the scroll in the middle of Herbert Hoover's speech, and Hoover had to slow down with it. "This damned thing, I could do better without it," he muttered, and the operator picked the tempo back up [21].

It's a recollection, from an oral history fifteen years later. But it may be the first recorded case of pacing to the machine.

So for most of the prompter's history, a person did the following. The fixed-speed scroll belongs to software prompters for people working alone, with nobody to turn the scroll. It solved the staffing problem, and it quietly handed the timing to the machine. There's a longer telling in the trouble with fixed-scroll teleprompters.

What we don’t know yet

No published study we could find compares a fixed-speed scroll with a voice-following one directly: the same speakers, the same script read both ways, and listeners judging the takes.

The closest is Trinity, a prompter for live presentations that follows the presenter's speech [22]. Its 62 presenters were rated higher on pitch, rate, volume and eye contact than with two baseline prompters.

What a prompter has to do

If timing is the problem, a prompter is only as good as what it does with your timing. Here are four things to look out for in any prompter tool:

1. **Waits when you pause.** Stop mid-paragraph for five seconds. The text shouldn't move. If it does, every thinking pause in a take is a pause you'll rush.
2. **Holds when you ad-lib.** Say a sentence that isn't in the script. The prompter should keep your place, not hunt for a match and jump. And it should pick up again when you're back on the script.
3. **Goes back when you restart a line.** Fluff a sentence, pause, and start it again. The prompter should move back with you, not stay where you stopped or leap ahead.
4. **Keeps the text near the lens.** The line you're reading should sit as close to the camera as the screen allows, and stay there. The gaze research above is the reason: direct gaze made listeners believe more and made viewers feel addressed [6] [16].

A fixed-speed scroll fails the first three by design. That isn't a fault in any one product. It's what a fixed speed means. A prompter with a hand control or a remote can pass them, as long as someone's working it.

What a prompter has to do

Four checks for any teleprompter. Each takes under a minute.

Waits when you pause

Stop for five seconds. Nothing moves.

Holds when you ad-lib

Say a line that isn't in the script. Your place stays put.

Goes back when you restart a line

Pause, start the sentence again. It moves back with you.

Keeps the text near the lens

The line you read sits right under the camera.

cuekards.com/blog/timing-not-the-script

How cuekards handles this

Cuekards was built around those four behaviours. Its Follow-my-voice feature listens as you read and moves the script with you.

- **It waits.** Silence never moves the script, and there's no time limit. Three seconds or five minutes, your place is there when you start again.

- **It holds when you ad-lib.** Words that aren't in the script nearby bring up an **OFF-SCRIPT** label, and your place stays put until you're back on a line it can see.
- **It goes back with you.** About a second of silence before a restart lets it follow you back, up to about a sentence. Further than that, press Enter on the desktop to re-sync.
- **The text stays near the lens.** The line you're on sits at a fixed reading line, and Text align moves that line to the top, middle or bottom of the screen to match where your camera is.

The browser version is free, in Chrome, Edge, Brave or another Chromium-based browser (version 120 or newer), on a desktop or laptop. For filming on a phone, there's the cuekards app for Android on Google Play, paid once with no subscription.

The limits. Follow my voice is currently available in English only, so a script in another language uses a constant-speed scroll instead. There's no iPhone or iPad app, and phone browsers aren't supported.

And it isn't flawless. A skipped paragraph or a phrase that repeats later in the script can still throw it, which is what the re-sync is for. How voice-following works goes through the failure cases one by one, and how it works covers the setup.

Why do I sound like I’m reading on a teleprompter?

Mostly timing. Listeners pick out read speech by its rate and its even, regular pauses more reliably than by its pitch [13] [11]. A fixed-speed scroll makes that worse, because it decides when you pause. The prompter settings that add to it are in how to read a teleprompter without sounding like you're reading.

Can people tell when you’re reading from a script?

Usually, yes. In lab tests, listeners tell read from spontaneous speech well above chance [11] [12]. With word-for-word identical scripts it gets much harder, and in one study listeners who hadn't done the task themselves were at chance [15].

Does a voice-following teleprompter make you sound more natural?

Nobody has tested that directly yet. What the research supports is narrower: timing is a large part of what makes speech sound read, and a following prompter leaves the timing to you. The test that would answer the question is described above.

Does looking at the camera make viewers trust you more?

On average, yes. Direct gaze made listeners believe more statements [6] and made video viewers feel more personally addressed [16]. The exception is a sceptical audience: viewers who already disagreed resisted more when the speaker looked at them [17].

The thing to try

Record the same 30 seconds twice. Once on a fixed scroll at your usual speed, and once with a prompter that waits for you: a friend scrolling by hand, or cuekards with Follow my voice ticked. Then listen to both without watching, and mark where each pause falls. In one take, the pauses will sit where the sentence needs them. That's the one to keep.

References

1. University of Wisconsin-Madison News (2000, 16 January). King of speakers: "I Have a Dream" was best of century (https://news.wisc.edu/king-of-speakers-i-have-a-dream-was-best-of-century/). The survey behind it was published as Lucas, S. E., & Medhurst, M. J. (2009), *Words of a Century: The Top 100 American Speeches, 1900-1999*, Oxford University Press.
2. Jay, R. (2013, 28 August). 50th anniversary of the March on Washington (https://www.tvobscurities.com/?p=14404). *Television Obscurities*. Draws on the *New York Times*' 1963 account of the networks' coverage.
3. The Martin Luther King, Jr. Research and Education Institute, Stanford University. "I Have a Dream" (https://kinginstitute.stanford.edu/i-have-dream).
4. Druckman, J. N. (2003). The power of television images: The first Kennedy-Nixon debate revisited (https://doi.org/10.1111/1468-2508.t01-1-00015). *The Journal of Politics*, 65(2), 559-571.

5. Rodero, E., Potter, R. F., & Prieto, P. (2017). Pitch range variations improve cognitive processing of audio messages (https://doi.org/10.1111/hcre.12109). *Human Communication Research*, 43(3), 397-413.
6. Kreysa, H., Kessler, L., & Schweinberger, S. R. (2016). Direct speaker gaze promotes trust in truth-ambiguous statements (https://doi.org/10.1371/journal.pone.0162291). *PLoS ONE*, 11(9), e0162291.
7. Van Zant, A. B., & Berger, J. (2020). How the voice persuades (https://doi.org/10.1037/pspi0000193). *Journal of Personality and Social Psychology*, 118(4), 661-682.
8. Guo, P. J., Kim, J., & Rubin, R. (2014). How video production affects student engagement: An empirical study of MOOC videos (https://doi.org/10.1145/2556325.2566239). *Proceedings of the First ACM Conference on Learning @ Scale*, 41-50.
9. Carpenter, S. K., Wilford, M. M., Kornell, N., & Mullaney, K. M. (2013). Appearances can be deceiving: Instructor fluency increases perceptions of learning without increasing actual learning (https://doi.org/10.3758/s13423-013-0442-z). *Psychonomic Bulletin & Review*, 20(6), 1350-1356.
10. Toftness, A. R., Carpenter, S. K., Geller, J., Lauber, S., Johnson, M., & Armstrong, P. I. (2018). Instructor fluency leads to higher confidence in learning, but not better learning (https://doi.org/10.1007/s11409-017-9175-0). *Metacognition and Learning*, 13(1), 1-14.
11. Dellwo, V., Leemann, A., & Kolly, M.-J. (2015). The recognition of read and spontaneous speech in local vernacular: The case of Zurich German (https://doi.org/10.1016/j.wocn.2014.10.011). *Journal of Phonetics*, 48, 13-28.
12. Blaauw, E. (1995). On the perceptual classification of spontaneous and read speech (https://www.coli.uni-saarland.de/Phonetics/icphs/ICPhS1995/13_ICPhS_1995_Vol_3/p13.3_254.pdf). *Proceedings of the 13th International Congress of Phonetic Sciences*, Vol. 3, 254-257.
13. Ryant, N., & Liberman, M. (2016). Automatic analysis of phonetic speech style dimensions (https://doi.org/10.21437/Interspeech.2016-1355). *Proceedings of Interspeech 2016*, 77-81.
14. Campione, E., & Véronis, J. (2002). A large-scale multilingual study of silent pause duration (https://doi.org/10.21437/SpeechProsody.2002-35). *Proceedings of Speech Prosody 2002*, 199-202.
15. Morris-Haynes, C., White, L., & Mattys, S. (2016). Listeners' discrimination of read and spontaneous speech is primed by performance of a prior speech production task

(https://www.isca-archive.org/speechprosody_2016/morrishaynes16_speechprosody.html). *Proceedings of Speech Prosody 2016*.

16. Kuang, Z., Wang, F., Xie, H., Mayer, R. E., & Hu, X. (2023). Effect of the instructor's eye gaze on student learning from video lectures: Evidence from two three-level meta-analyses (https://doi.org/10.1007/s10648-023-09820-7). *Educational Psychology Review*, 35(4), 109.
17. Chen, F. S., Minson, J. A., Schöne, M., & Heinrichs, M. (2013). In the eye of the beholder: Eye contact increases resistance to persuasion (https://doi.org/10.1177/0956797613491968). *Psychological Science*, 24(11), 2254-2261.
18. Associated Press (2000, 21 December). Barney McNulty, creator of cue cards, dies at 77 (https://www.deseret.com/2000/12/21/19545505/barney-mcnulty-creator-of-cue-cards-dies-at-77/). *Deseret News*.
19. Cohen, E. (2000-01, Winter). A figure of speech (https://magazine.nd.edu/news/17699). *Notre Dame Magazine*.
20. Mechanical memorizers: TV stars rely on teleprompters to help remember their lines (https://archive.org/stream/tvguide-chicago-1953-12-04/tvguide-chicago-1953-12-04_djvu.txt) (1953, 4 December). *TV Guide* (Chicago edition).
21. Schwartz, T. (2016, 8 June). Hoover and the teleprompter (https://hoover.blogs.archives.gov/2016/06/08/hoover-and-the-teleprompter). *Hoover Heads*, Herbert Hoover Presidential Library, National Archives.
22. Wu, Y., Li, S., Zhang, S., Wang, X., & Li, Q. (2024). Trinity: Synchronizing verbal, nonverbal, and visual channels to support academic oral presentation delivery (https://arxiv.org/abs/2411.17015). *Proceedings of Chinese CHI 2024*.