

## COMMENTARY: LEARNER-BASED LISTENING AND TECHNOLOGICAL AUTHENTICITY

**Richard Robin**

**George Washington University**

*You dropped a hundred and fifty grand on an education you could've picked up for a dollar fifty in late charges at the public library.*

Good Will Hunting, 1997

Language teachers know that even the best technology cannot provide the high degree of interaction required to acquire meaningful proficiency in a foreign language. Even the most polished packages available today (and likely to be available for several years to come) cannot evaluate learner input and provide subtle shades of context-based feedback, except in the narrowest of circumstances. Technology's dull blade is even more apparent the moment interactive orality is required. A simple phone call to a voice-automated service center reminds us to what extent mass-market speech recognition is crude and speaker-finicky, even in *English*. The speakers of less common languages may have to wait years before work begins on speech recognition for speakers of their languages.

Off-the-shelf technology is not ready for interactive oral-aural instruction, but it has reached a level of sophistication that makes it ideal for use by the strategically independent learner to acquire and improve receptive skills *in an authentic environment*, if we update our definition of authenticity to include the technologically-enabled possibilities supporting a text or script<sup>1</sup>: the availability of combined texts and scripts, user-control over script delivery both in terms of speed and chunking, user-created glossing aides, captioning, etc. This technological overlay, available not just to language learners but to all users, is "authenticizing" practices that were once considered inauthentic. No longer are such devices part of the specialized landscape of the L2 learner; instead they make up the everyday L1 machine-mediated world of listening. That has implications for the demands learners make of themselves and the tasks that they choose. It also leads us to reexamine the value of pre-packaged listening comprehension materials in which L2 listeners are guided in listening strategies but are not encouraged to make use of technological innovations that native listeners are coming to use on a regular basis.

A brief survey of the available user-directed modifications to online scripts leads one to the idea that in the immediate future — the next five to ten years — the frontier in language learning and technology will not be found in what program does what better, but rather which students use off-the-shelf technology to best facilitate their own learning in their own learning style. Just as we began to teach metacognitive acquisition strategies, such as the use of background knowledge and prediction in the 1980s (see Nunan, 1999; Omaggio-Hadley, 2001; and Ur, 1984) for summaries of research and practice then and now), we should now teach meta-technical skills to language learners, rather than setting them out on a closed loop. Others have come to the same conclusion that technological literacy is an essential component of the language acquisition strategic toolbox (Godwin-Jones, 2000; Hubbard, 2004; LeLoup & Ponterio, 2000; Richards, 2000;). Effective users of raw electronic resources, such as easily repeatable video clips, captions, and even translation bots will bring a wider variety of input at the proper level for a broader range of learning styles than could possibly be made available in any pre-packaged closed-track program.

In listening comprehension, attention over the last twenty-five years has turned from adapted scripts to authentic audio and video in which we scaffold "real media" with remediated tasks (pre- and post-script activities) instead of changing the media itself. Yet despite the scaffolding, authentic materials, particularly where they involve a non-interactive flow of speech such as radio, TV, and movies, remain a challenge, especially in the beginning stages of language learning. The audio is too fast. Or acoustically

difficult. Or too heavily culturally referenced. Or has too much slang. Take the scaffolding away, and the learner's activity falls apart.

As a result, materials designers continued to concentrate on wrapping the materials in better wrappers, e.g., by adjusting the task and occasionally modifying the script (we call it *semi-authentic*) rather than teaching learners to use the technology to mediate the script. Such sites abound on the web. Many, such as [SCOLA](#), rely on authentic video and audio with wraparound exercises and transcripts. Others such as [Randall's ESL Cyber Listening Lab](#) or the [NCLRC's Simplified Russian Radio Site](#) (created by the author of this article) resort to semi-authentic audio.

On the other hand, commonly available technological fixes can be called upon for greater degrees of remediation. At one time, such modifications were considered distinct from "authentic" ([Chapelle, 1998](#)), but today they all part of the world of the native listener that involves neither scaffolding nor semi-authenticity. Here are some of the more prominent devices.

### Repeated audio delivery

Before 2000, most broadcast media fare was not by default user-repeatable. Now only legacy media, i.e., analog television and radio broadcasts of over-the-air origin, remain in that category. Digital video recorders (DVRs), which can simultaneously record and playback broadcast television, have begun to change the basic psychology of frame-by-frame repeatability. Cable companies have joined the recording revolution by offering DVR cable boxes. In computer-mediated video and audio, repeatability predominates. Progressively downloaded clips are by default repeatable. Streaming media can usually be captured and repeated, although this is increasingly subject to various digital rights management (DRM) schemes and hastily found workarounds. The notion that L2 learners must grab a flow of speech on the first try or lose the meaning is valid only for those events where the audio is not repeatable. For electronic media, that is fast becoming the minority of situations. In short, *listening has become a semi-recursive activity, less dependent on transient memory*<sup>2</sup>, inching its way closer to reading, which is fully recursive. In fact, the presentation of authentically repeated audio scripts has left the world of cable TV and the desktop computer and invaded more mundane areas such as phone service call menus ("Press 9 to repeat these options."), answering machine messages, recorded live events on iPods, and so on.

### Slowed audio text delivery

Fast delivery rates intimidate listeners and impede L2 comprehension. There is evidence that globalization has resulted in increased delivery tempos in areas of the world where Western broadcasting styles have replaced traditional authoritative styles. In the former Soviet Union, for example, delivery as measured in syllables per second has nearly doubled since the fall of communism from three to about six syllables per second (Robin, 1991). Casual comparisons for news broadcasts in places such as the Middle East and China lead to similar conclusions. *Semi-authentic* projects, such as the [NCLRC's Simplified News in Russian \(Robin and Bessergeneva\)](#) give listeners a slower newscast together with a transcript and scaffolding exercises. However, enterprising learners need not depend on such semi-authentic content. Instead they can take advantage of any number of truly authentic online recorded newscasts, complete with transcripts and background information. The trick is slowing down the stream of language. That's possible for most of common file types and players (among them divx, mp3, mp4, wmv, wma, and mov in Windows Media Player and QuickTime, as well as in many generic players). As long as the source file diction will withstand slowdown, i.e., the script reader is not sacrificing syllables to attain speed, language learners can avail themselves of a wide variety of authentic material. Using more sophisticated editing software, such as [Sourceforge's Audacity editor](#), users can not only slow down speech but also insert pauses.

### Accompanying texts

Those who pioneered pedagogical approaches to non-interactive listening comprehension in the late 1970s and 1980s avoided the use of full-text supporting scripts, preferring instead semi-script outlines (Geddes & White, 1978) and various other round-about scaffolding devices such as true-false items, multiple choice, direct-question, cloze exercises, and fill-in-the-map (Chauvin, 1980; Meyer, 1984; Ur, 1984, p. 77-85). But today news and public affairs websites often accompany their online webcasts with transcripts or summaries of the report as broadcast or webcast. Even when the transcripts are missing, a backgrounding text of similar content is only a few mouse clicks away. In such situations, the learner, just like the native listener, is in control of how much additional information is needed. Of course, the availability of additional information is of no use unless learners become aware of when they require it for more complete comprehension of the script at hand.

### Captioned video

For many years a widespread view on audio comprehension held that both target-language captions and native-language subtitles were anathema to developing listening comprehension. But this popular view has not been well tested. The research consensus (Garza, 1991; Hwang, 2004; Jones, 2003; Markham, 2000-2001; Park, 2004; Stewart & Pertusa, 2004) suggests that L2 captions aid in immediate comprehension (hardly an earth-shattering finding). But we know little about the longitudinal effects on learning in terms of listening comprehension improvement or retention of incidentally acquired vocabulary, either in receptive or productive modalities. Yet in the last ten years the availability of closed-caption video has mushroomed, both on DVD and in a number of countries through ordinary analog broadcasts. Moreover, even versions of video that come un-subtitled are readily captioned. Simple text-format SubRip Title (SRT) scripts (see Figure 1), available at Internet subtitle databases such as [opensubtitles.org](http://opensubtitles.org) combined with free caption-reading players (among them Windows Media Classic, [Z-player](#), and [VLC](#)), make do-it-yourself closed-captioning available to any viewer with a web connection. The simplicity of text-only SRT scripting makes it possible for teachers with the technological wherewithal to record video locally and caption for the target audience: L1 where required, L2, when appropriate, and laconic L2 glosses where possible.


 <p>Screen shot of Line 2 of SRT script</p>	<pre> 1 0:01:20.500 --0:1:23.000 Ты кого-нибудь когда-нибудь убил?  2 0:01:23.500 --0:1:25.000 А разве это плохо? really </pre> <p>SRT script of Lines 1 and 2 in the clip. This is the complete file. There is no additional HTML or SMIL markup. Most generic players can be configured to read SRT files of the same name in the same directory by default.</p>
<p>English translation:  Line 1 – Have you ever killed anyone?  Line 2 – Is that really something bad?</p>	

Figure 1. Example of targeted captioning of soap opera segment for students of Russian

### Translation bots and listening comprehension

Bots automatically generate text, based on user input. They have already attracted attention in language production practice. For example, [Fryer and Carpenter \(2006\)](#) suggest that language learners use chat bots for practice. In listening comprehension, one can imagine scenarios where learners could make use of primitive Web-based translation bots, such as [Google's Language Tools](#) or [Alta Vista's Babelfish](#) in a way that would redefine the notion of usable learner input.

Until recently, teachers and learners could expect limited use of most connected paragraphed scripts, i.e., news, weather, advertisements, etc., at the elementary level. At best, beginning learners could expect to determine topic and predict the general content by picking out a few common words and some cognates from an unfamiliar script. Traditionally we gave such activities, emphasizing less than working comprehension, two labels: (1) fitting the task, not the text, to the learner's level and (2) teaching the learner to live with the frustration of not understanding everything. Those limitations were forced upon us by linguistic reality and technological limitations. However, with the advent of internet-based machine translations of text, beginning learners can tackle the same elementary tasks in a far more satisfying way. They can find paragraphed audio with accompanying online texts. They can put the texts through a translation bot into L1. The result is good enough to provide the key words, the gist and main details. With a basic understanding of what will now be said, they can listen to the audio, using all the enriched background at their command to make out more of the language. Such practices reverse the bottom-up approach to listening forced on users by the linearity of audio alone. Of course, to a large extent a learner's style is likely to determine how much top-down strategies will predominate. The point is that the technology now available to all makes that choice a reality.

### Voice chat and interactive native speaker practice

identifies synchronous voice chat programs such as [Skype](#) as "disruptive" technologies: those which change the face of an industry (in this case phone and conference service). Voice chat, which is free, increases the availability of practice with native speakers. The practice of e-mail exchanges is now being joined by audio chat exchanges ([Volle, 2005](#)). We usually think of audio chat in terms of the development of speaking proficiency. However, extensive audio chat involves greater information gaps than face-to-face practice with native speakers. Consider the increase in effort for interactive listening comprehension, required by these additional factors:

- 1) In face-to-face exchanges, learner and interlocutor share the same physical surroundings, which eases mutual comprehension when talking about oneself and daily activities. In audio chat, the learner and native interlocutor may be thousands of kilometers apart with no shared local space or shared day-to-day realia — reference to local practices or events ("I have to cut our chat short; the construction on the Whitehurst Freeway is awful and I have to get home in time.")
- 2) In face-to-face exchanges taking place in the learner's L1 country, the target-language native interlocutor is likely to have some proficiency in the learner's L1 and might be tempted to translate when comprehension fails.
- 3) In local exchanges the native interlocutor is likely to have had more experience with the learner-type (e.g. a Russian in an American university is likely to have spoken to other American students and has an idea of what American university students can say and understand in Russian). In distance exchanges, at least initially, both learner and interlocutor are likely to know less about each other's language repertoire. That would require more in interactive meaning negotiation.
- 4) Face-to-face exchanges allow participants to resort to facial expressions and gestures. Audio chat does not.

In the spirit of *Good Will Hunting* quoted at the beginning of this piece, one could argue without much difficulty that the technology of 1990 made self-directed efforts at L2 reading acquisition an easy matter

for learners who had mastered basic learning strategies and had reliable Internet access. Now, seventeen years later, the technological infrastructure for listening allows learners to do the same, but with an important difference: we have seen that the technology has in fact changed the definition of natively authentic listening tasks. Those pursuing the listening skill now have a wider range of engaging scripts to work with at a variety of levels. That in turn provides the additional time on task required for listening. Those who are both strategically and technologically prepared can direct their own learning, adjusting the process for their own styles and goals, all within the bounds of an experience that approaches the natively authentic and with less reliance on artificial scaffolding.

The rapid spread of technology (e.g. broadband connections, webcams, online communications, and collaborative software), along with a trickle-down technological proficiency from enthusiasts to everyday users, promises to broaden the notion of using the "raw" electronic world unmediated by pedagogical middlemen and is applicable to modalities beyond listening. It takes little to dream up online self-directed activities such as mixed oral written chat groups, native spell-check grammar editing using both machine and live editing, or, even farther in the future, truly accurate self-directed phonetics practice. For technologically proficient users who are also metacognitively aware, the horizons of task-based language learning are wide indeed.

The challenge to the foreign language *teacher* is daunting but clear. Of those instructors who actively use technology, many control the tools at hand only to perform well-rehearsed procedures that have been laid out for them in detail. Few are in a position to advise their students on how to use the technology as a language learning enabler. Fewer still can quickly envision uses for a new "disruptive" technology. The challenge then is to enable our teachers to enable our students. And with such a wide range of technological prowess among the ranks of our profession, that is no small task. We should not, however, refrain from asking those entering the profession to give ever more thought to what we can do with off-the-shelf technology beyond ready-made pedagogical packages.

---

## NOTES

1. For the purposes of this paper a text is written, while a script is delivered orally.
2. Transient memory, ranging from immediate "echoic" memory for less meaningful utterances to primary memory with durations between a few seconds and 30 seconds of information (Cowan, 1995, p. 338; Miller & Johnson-Laird, 1976, p. 144).

---

## ABOUT THE AUTHOR

Richard Robin, Associate Professor and Language Program Director for Russian at George Washington University, received his Ph.D. in Slavic linguistics from the University of Michigan and has been at George Washington since 1981. He also serves as the Language Center's Technology Specialist. His main area is methodology of Russian language teaching and technology in language teaching. He additionally coordinates distance-learning projects using authentic foreign-language materials on the Internet and serves as a senior researcher at the [National Capital Language Resource Center](#).

E-mail: [rrobin@gwu.edu](mailto:rrobin@gwu.edu).

**REFERENCES**

- Chapelle, C. A. (1998). Multimedia CALL: Lessons to be learned from research on instructed SLA. *Language Learning & Technology*, 2(1), 22-34.
- Chauvin, J. (1980). Exploitation d'une emission radiophonique. *Le français dans le monde*, 158, 64-67.
- Cowan, N. (1995). *Attention and memory: An integrated framework*. New York: Oxford University Press.
- Fryer, L., & Carpenter, R. (2006). Emerging technologies: Bots as language learning tools." *Language Learning & Technology*, 10(3), 8-14.
- Garza, T. (1991). Evaluating the use of captioned video materials in advanced foreign language learning. *Foreign Language Annals* 24(3), 239-258.
- Geddes, M., & White, R. (1978). The use of semi-scripted simulated authentic speech in listening comprehension. *Audio-Visual Language Journal*, 16(3), 137-145.
- Godwin-Jones, R. (2000). Emerging technologies: Literacies and technology tools/trends. *Language Learning & Technology*, 4(2), 11-18.
- Godwin-Jones, R. (2005). Skype and podcasting: Disruptive technologies for language learning. *Language Learning & Technology*, 9(3), 9-12.
- Hubbard, P. (2004). The challenge of learner training for CALL. In S. Fotos & C. M. Browne (Eds.), *New perspectives on CALL for second language classrooms* (Chapter 6). Mahwah, NJ: Lawrence Erlbaum Associates.
- Hwang, Y-L. (2004). The effect of the use of videos captioning on english as a foreign language (efl) on college students' language learning in Taiwan. Doctoral dissertation. Ann Arbor, MI: UMI.
- Jones, L. C. (2003). Supporting listening comprehension and vocabulary acquisition with multimedia annotations: the students' voice. *CALICO Journal*, 21(1), 41-65.
- LeLoup, J. W., & Ponterio, R. (2000). Literacy: reading on the net. *Language Learning & Technology* Vol. 4(2), 5-10.
- Markham, P. (2000-2001). The influence of culture-specific background knowledge and captions on second language comprehension. *Journal of Educational Technology Systems*, 331-343.
- Meyer, R. (1984). 'Listen my children, and you shall hear....' *Foreign Language Annals* 19(3), 203-208.
- Miller, G. A., & Johnson-Laird, P.N. (1976). *Language and Perception*. Cambridge, MA: Belknap / Harvard University Press.
- Nunan, D. (1999). *Second language teaching and learning*. Boston, MA: Heinle and Heinle.
- Omaggio-Hadley, A. (2001). *Teaching language in context*. Boston: Heinle & Heinle.
- Park, M. (2004). The effects of partial captions on Korean EFL learners' listening comprehension. Doctoral dissertation. Ann Arbor, MI: UMI.
- Richards, C. (2000). Hypermedia, internet communication, and the challenge of redefining literacy in the electronic age. *Language Learning & Technology*, 4(2), 59-77.
- Robin, R. (1991). Russian-language listening comprehension: where are we going? where do we go? *Slavic and East European Journal*, 35(3), 403-410.
- Stewart, M. A., & Pertusa, I. (2004). Gains to foreign language while viewing target language closed-caption films. *Foreign Language Annals*, 37(3), 438-443.

Ur, P. (1984). *Teaching listening comprehension*. Cambridge: Cambridge University Press.

Volle, L. M. (2005). Analyzing oral skills in voice e-mail and online interviews. *Language Learning & Technology* 9(3), 146-163.