



HAL
open science

(M)other Tongue

Denyze Toffoli, Ana Bumber, Opeyemi Oso

► **To cite this version:**

| Denyze Toffoli, Ana Bumber, Opeyemi Oso. (M)other Tongue. 2024. hal-04779544v2

HAL Id: hal-04779544

<https://hal.science/hal-04779544v2>

Preprint submitted on 26 Nov 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial - NoDerivatives 4.0 International License

(M)other Tongue: Towards generating an ideal L2 self

Denyze Toffoli, Ana Bumber and Opeyemi Oso

Laboratoire Interuniversitaire de Recherche en Didactique des Langues (Lairdil), Université Toulouse III – Paul Sabatier, Toulouse, France

Author biographies and affiliations:

Denyze Toffoli is a full professor whose research interests concern primarily the online informal learning of English, often extended to all foreign language development outside the classroom. She favours psycholinguistic perspectives on learning, but is also concerned with how classroom interaction can support extramural learning. Her supervision of recent theses and the mentoring of new colleagues has led to her recent interest in AI applications for language learning.

Ana Bumber holds a PhD in literature (Nabokov) from the University of Tours in France. Since her arrival at the University of Toulouse III in 2022, she has been teaching English for Specific Purposes to students of Chemical and Process Engineering and has invested in research devoted to language teaching and didactics, specifically concerning the use of different AI tools, such as automatic speech recognition, automatic captioning, and video translators.

Opeyemi Oso is a doctoral student in Applied Linguistics. Her thesis concerns digital technologies for teaching and learning L2 French & English pronunciation.

The authors are all members of Lairdil (*Laboratoire Interuniversitaire de Didactique des Langues*), which focuses on research in the learning and teaching of languages for specialists of non-linguistic disciplines, at the Paul Sabatier University in Toulouse. They are all involved in projects dedicated to studies of AI usage for language learning in formal and informal contexts.

Acknowledgements

We would like to thank all of our colleagues at the Université Toulouse III – Paul Sabatier who participated in this study, both evaluating the langue produced by the AI and lending themselves to experimentation with the videos.

Statements on open data, ethics and conflict of interest

- a. All data is stored on the French National *Huma-Num* platform, can be accessed by request to the authors.
- b. The study follows ethical guidelines as provided by the 2023 Revised Edition of *The European Code of Conduct for Research Integrity* and BERA's 2018 Fourth Edition of *Ethical Guidelines for Educational Research*. All participants in this study signed a consent form agreeing to the procedures followed and that data treatment proposed. After collection, all data was anonymised with random reference numbers and correspondence sheets were destroyed.
- c. There are no conflicts of interest between the authors, participants and persons involved with the commercial products used for this study.

(M)other Tongue: Towards generating an ideal L2 self

The *(M)other Tongue* project is a contribution to research in applied linguistics on the use of intelligent tools (AI) for additional language learning. 2023 saw the commercialisation on-line of several AI-enabled multilingual video services, with at least one including lip-synchronisation and translation functions. The *(M)other Tongue* project aims to determine relevant pedagogical uses of this type of video-platform for foreign language learning by evaluating the possible perceptive and imitational value of translated videos of self with adult learners of additional languages. This preliminary study worked with fifteen volunteer university lecturers whose videos of themselves speaking their L1 were translated into another language. Both the quality of the video products and the emotional impact on the participants were then assessed and recommendations made as to how such tools could further several areas of applied linguistics inquiry, including ideal L2 self theory, motivation, imitation and pronunciation. This research note presents and discusses this preliminary study and the next steps envisaged for a larger-scale study with non-specialist students of English.

Keywords: AI-generated video translation, ideal L2 self, imitation, visualisation

Introduction

Technological invention accompanied by widespread access to new tools have always driven Computer Assisted Language Learning (CALL) and the arrival of generative AI (GenAI) is no exception. In 2023 a number of online apps integrated video-generation features in multiple languages. One of these in particular professes to “effortlessly translate your videos using [a] one-click solution that clones your natural speaking voice and style for seamless delivery in other languages” (<https://www.heygen.com/video-translate>). This function initially caused us, the authors, some concern, similar to that provoked in our field by ChatGPT3.5 (Alm & Watanabe, 2023; Kohnke, Moorhouse, & Zou, 2023), as we regularly evaluate our students’ English via video presentations that we ask them to create and upload to our learning management platform (LMS). However, we decided to take a more constructive approach and study the potential of this type of video adaptation for language learning. In

reference to Dörnyei's (2009) concept of the ideal L2 self, our research questions became: "How could observing oneself speaking an additional language (Lx) fluently effect one's ideal Lx self-image? What might the motivational effects of this fluent self-image be? Could this "fluent self" provide a model for posture, pronunciation (specifically articulatory movement) or other language-related gestures? This article covers the theoretical frameworks and preliminary phase of this research, which involved the participation of 15 foreign language teachers and researchers in France. Each participant recorded two 30-second videos in their L1 (which generally corresponded to French or English, but also included Croatian, Arabic and Spanish) and answered a short questionnaire concerning their language preferences for translations. The videos were then translated and transposed using a commercial online video-creation site into the participant's choice of two other languages (from among Arabic, Croatian, English, French, German, Hebrew, Italian, Russian, Spanish and Polish). The videos thus produced were both assessed by a native speaker, to evaluate their linguistic quality and the vocal and lip-synching realism, and made available to the participants themselves. The participants' reactions to seeing themselves speaking their Lx fluently were then collected via an online questionnaire. As they are all teachers, they were also asked to communicate their thoughts on possible pedagogical uses for such products. These results are presented herein.

Background, Theoretical Frameworks and Literature Review

Widescale public access to generative AI products has seen the creation of AI-powered sites, such as *FlexClip*, *Fliki* or *Pictory* that generate video content from text or audio files. Others, like *Steve.ai* or *Synthesia* customize accents, sometimes generating avatars. *SyncLabs* and *Synthesia* offer lip-synching, which allow the avatars or real-person videos to look realistic when pronouncing words from text. Most of these apps can deal with multilingual content. Some translate audio files from one language to another or generate audio in one language from text in another. Among those that present text-to-video creation in different languages or

with the capacity to create video in a different language from the text entered are *Predis.ai* and *FlexClip*. *ElevenLabs* provides tools for text-to-speech, speech-to-speech, and audio dubbing translation with voice generation or voice preservation, the latter allowing the original speakers' vocal identity to be heard. At the date of this writing (November 2024), it offers speech options in 29 languages and includes a "voice-cloning" capability. *HeyGen* however appears to be the only app at the moment that not only uses audio translation and voice-cloning, but also synchronizes lip movements, to produce realistic visuals including articulatory movement. This function is advertised as localisation for marketing or other business videos, but it would seem to have particular potential for language learning.

Films and other video material (series, podcasts and diverse video-clips) have long been part of the language learner's toolkit and have also been widely used as teaching aids in the foreign language classroom (Léon-Henri, 2012; Rémon, 2012; Vanderplank, 2016). One reason for this is that the literature in applied linguistics has provided a general consensus that extensive exposure to a target language contributes significantly to its acquisition (N. C. Ellis, 2002; R. Ellis, 2012; Hilton, 2014;), going back at least to Stephen Krashen's work on his input hypothesis (Krashen, 1985). Video has also been found to be engaging, involving both visual and auditory stimuli, two of the most important senses for language, simultaneously.

Cognitive Load Theory (Sweller, Ayers & Kalyuga, 2011) has established the importance of multimodality in education and the diverse (sometimes positive and sometimes negative) effects of material presented visually and auditorily at the same time. Their "redundancy effect" (Sweller et al., 2011b) states that coherent redundancy between visual and auditory cues can be particularly helpful for novices. A study by Inceoglu (2016) indicated that while both audio-only and audio-visual training helped L2 learners of French with their perception of nasal vowels, audio-visual training produced superior, statistically significant results in their production. Vanderplank (1988, 1990, 1993, 1996) and Halliday

(1989) have argued that bi-modal input (image and video) is beneficial as it takes learners' different learning styles into account. Other empirical research has consistently shown that experimental groups viewing with captions outperform groups with no captions (Huang & Eskey, 1999; Montero Perez, Peter, Clarebout & Desmet, 2014; Rodgers, 2013b), showing captioned viewing to enhance both vocabulary learning and video comprehension. Cognitive Load Theory does caution however that redundancy may be distracting, or even counterproductive, for experts. This could be the case, for example, with subtitles or closed captions for bilinguals or highly proficient Lx users.

In psychology, self-image is recognised as an important factor for learning. A positive self-image is seen as motivating and negative self-image as demotivating. In the field of applied linguistics, Zoltan Dörnyei has established the notion of the 'ideal L2 self' (Dörnyei, 2009; Dörnyei & Ushioda, 2009), which is said to be a determining factor in engagement in, and even success at, language learning. The ideal L2 self is the best version of the self that could be imagined speaking another language. Dörnyei (2009) also explores the "ought to L2 self" which is a future self that is imposed by family or culture. He argues that by creating a positive vision of oneself using the Lx, one can positively influence language learning outcomes.

Nonetheless, relatively little attention has been paid to the power of imagination in SLA (Ryan & Irie, 2014). In the same way that positive visualisation has been used to improve performance in sports, Ryan & Irie (2014) and Dörnyei & Ushioda (2009) believe that it can be implemented in L2 learning. Toffoli & Allan (1989), for example, successfully used creative visualisation with primary school ESL immigrants to Canada in the 1980s. Dörnyei contends that the vividness of the visualisation of the ideal L2 self, as well as the plausibility of the image are two essential factors in its motivational potential (2009). Ryan and Irie (2014) discuss how mental images of the self can modify attributions (of success or

failure), thus influencing motivation. They specify that process-oriented mental simulation is more effective than goal-oriented simulation (p.115) and that third-person visualisations are generally more compelling than first-person perspectives (p.116).

Video and audio modification can provide external means of directly accessing representations of the self, with the potential to bypass difficulties that mental visualisation may encounter. They also effectively create a third-person perspective, as the self on screen is physically separate from the self as it is experienced internally. A concrete video image of a learner speaking their target language fluently could therefore be a potent means to providing a strong, positive Lx self-image, perhaps particularly for those who have a poor perception of their Lx competence or of their ability to adequately learn their target language.

Imitation theory stresses the importance of mimicry in phonological production (Markham, 1997; Nguyen & Delvaux, 2015) and many language learning methodologies are based on imitation and repetition (Cuq & Gruca, 2005; Duff, 2000), not only for pronunciation, but also for syntax or vocabulary acquisition. However, in this type of approach, the speaker to be imitated is often a 'native' speaker, i.e. someone speaking their mother tongue, or a recognized expert, such as a teacher. Several studies examined by de Meo, Vitale, Pettorino, Cutugno & Origlia (2013) have shown that close matches between the model's and the learner's voices produce better pronunciation achievement in listen-and-repeat exercises, calling closely matched models "golden speakers" and suggesting that "the most effective golden speaker to learn segmental and suprasegmental features of a second language is the learner's own voice with a native accent" (de Meo et al., 2013, pp. 90–91).

Taking oneself as a model could thus potentially have a strong positive influence on the way in which a learner perceives and then produces Lx speech. On this basis, Henderson & Skarnitzl (2022) modified the audio recordings of four university lecturers learning English to conform more closely to standard English intonation patterns and asked the learners to

listen and repeat, using their own modified recordings as a model. Three of the four learners found the activity helpful and a positive reinforcement for the “better me” that they aspired to be as speakers of English.

The complex interplay of this theoretical background combined with emerging video-translation technologies suggests new affordances for language learning and prompted the following research questions:

- Does seeing oneself speaking a foreign language of choice fluently induce changes in self-perception, in one’s image of the "ideal L2 self", in motivation or in other affective states related to learning that language?
- Does taking oneself as a model influence the way in which a learner perceives and produces the foreign language?

However, to be in a position to answer these questions appropriately, using accessible, online, AI-generated video-translation technology, it was necessary to carry out a preliminary study, to test the technology, the procedures to be put in place and resolve any ethical questions that the use of this technology might pose.

Therefore, while keeping our main research questions in mind, this study focuses on more pragmatic questions, to help prepare the main phase of the study. We chose to work with colleagues, targeting validation of the aforementioned research questions, exploration of methodological and ethical issues, identification of potential problems and collection of additional ideas on possible pedagogical uses, before carrying out further studies and concrete pedagogical intervention with students themselves.

Methodology

For this preliminary study we solicited the participation of colleagues in a French university during a pedagogical exchange day that took place in November 2023. After watching a short

presentation by the authors and a demonstration of the video of one of them (a Nigerian) fluently speaking the native language of another (Croatian¹), sixteen volunteers signed informed consent forms (see appendix 1) to take part in the study. One later withdrew. The only criteria for selection at this stage were to be teaching in higher education, to be interested in language learning and to agree to volunteer. The relatively small number of participants was considered sufficient to validate the use of the tools and the procedures of the intended protocol. It would also allow for quick qualitative treatment of the data collected.

All participants received a copy of the protocol, requesting them to film themselves (using their cell phone) in their native or primary language, first reading a (30-second) passage by a favourite author and second speaking spontaneously for 30 seconds about their hometown or about another topic that would not require preparation. They were requested to upload their video to a university-secured storage space (Renater – *Filesender*) and to answer 3 online questions to determine the target language(s) for translation.

The researchers then recovered the original videos, uploaded them to a video-translation platform (*HeyGen*) and had them translated/transposed into the desired language(s). The new videos were made available on storage space provided by *HumaNum*, a French platform for the secure management of research data in the Human and Social Sciences, and a link leading to their own video(s) was sent to each individual participant. They were also sent a link to a second online questionnaire with six open-question themes:

- (1) After watching the translation of your video, what is your reaction? What do you think when seeing yourself speaking this other language? How do you feel?

¹ Though the selected language was labelled “Croatian” and not “Serbian”, “Serbo-Croatian”, “Croato-Serbian”, “BCS” or “BCMS”, the accent and the manner of speaking was of the Eastern variant as spoken in parts of Bosnia or in Serbia.

- (2) Does the video seem realistic? Does it motivate you and if so how?
- (3) If possible, could you show your video to a friend or family member and report their reaction, stating whether or not the person speaks the language in question?
- (4) Do you think you could use this video to help you learn the language?
- (5) What pedagogical uses do you see for these types of video translations?
- (6) Please indicate your name if you agree to being identified.

Answers to the questionnaire were collected and analysed for content only. Other colleagues and personal contacts were also solicited to provide assessments of the language produced in the newly translated videos. The results are described below.

Results

Perceived quality of the generated videos

We turned to eight native speakers among our colleagues and personal contacts to provide us with spontaneous reactions to the quality of the language encountered in the generated videos. Each of them viewed all of the videos produced in their native tongue. A representative selection of their comments appears below.

English (7 speakers):

The accent is very English, although other accents could have been chosen. The speakers appear very serious and the articulations sound a tiny bit stilted, but the wording is natural, fluent and proficient. They would definitely get good marks on a spoken English exam.

Arabic (4 speakers):

The pronunciation is clear and precise, sounding natural and realistic. There were only two words across all the videos that were not pronounced correctly as they are in Arabic, but they are still easily recognisable. Also, the places of articulation in the mouth closely resemble Arabic pronunciation.

Russian (4 speakers)

This AI application is really very effective! For the most part, it sounds like native speakers (at least in terms of pronunciation). There are, however, a few inaccuracies in intonation and translation (especially in spontaneous speech).

As can be seen by reading through this selection of spontaneous reactions, the native-speaker evaluators found the translated videos convincing. In spite of a few imprecisions, both the (translated) language and voice cloning produced by this AI were judged to be accurate and authentic-sounding. In one instance the evaluator for Spanish declared: “I even recognise his voice! Crazy!²”

Perceptions of self as seen in the generated videos

Participants were asked to respond to our questionnaire immediately upon viewing the generated video. The first (open) question asked for their thoughts and feelings upon seeing themselves speaking another language. The most frequent adjectives used were “impress*” or “surprise*”³ (six occurrences each), followed by “amaz*”⁴ (three occurrences) and a series of single-use adjectives: extraordinary, fascinating, incredible, amusing, nifty (*chouette*), funny. Other than feeling impressed or surprised, respondents also felt “moved” or “excited” and expressed “admiration”, “fun”, and simply “I like it”.

While most of the participants had positive reactions to seeing themselves speaking a foreign language, a few negative reactions were also expressed, the speakers finding the translations “less natural”, “robotic”, “strange”, producing an effect of “otherness”, of “cognitive dissonance”, or in some way “not perfect”. These included feelings of not really

² je reconnais sa voix ! C’est fou !

³ We indicate the lexical root, as the forms varied depending on grammatical context.

⁴ « bluffé » or « bluffant » in French

recognizing oneself, saying that the generated video produced “neither my regular voice nor my regular accent” (this for a person who tested the translation on an Lx that she speaks very well) or that “I did not feel the emotions” associated with the original discourse.

Using the generated videos for learning an Lx

Regarding the respondents' potential personal use of AI translated videos for learning, seven responses mentioned focusing on pronunciation training, particularly segmental and suprasegmental elements. Respondents also mentioned using it to help learn grammar (“locating logical connectors, ... grammatical tenses”⁵), vocabulary (“choosing my own script could help me to learn vocabulary”⁶), and to improve listening (“I tried to identify words by listening to it”⁷). One respondent indicated the possibility of comparing the structure and forms of sentences (syntax) in the target language with that of languages already mastered. Her idea was to watch her own recording in her target language and, using subtitles in her own language, to compare and work through the differences. Another thought that the video could serve as a model for practicing through repetition and one felt that it could help with self-confidence. Thus most of the participants thought it would be possible to use such AI translated material to help them in the learning of their Lx, regardless of their present level of (non)expertise, although one participant did express reticence about its usefulness at the extremes of the learning continuum, stating that it would not really be very useful for her, as she already spoke English very well and not a word of Polish (the two languages she had chosen for the translations).

⁵ Repérage des connecteurs logiques, ... les temps grammaticaux

⁶ le fait de choisir mon propre script pourrait m'aider [à] ... enrichir mon vocabulaire

⁷ J'ai essayé d'identifier des mots en l'écouter

Using generated videos in teaching

Participants were asked how they might use AI-generated video translations as a teacher. Eleven of them mentioned pronunciation training, including prosody, intonation, imitation, articulation, and accent. Other pedagogical uses mentioned included practicing for oral presentations, although for one participant the tool was also perceived as a potential threat, allowing students to cheat if they had to submit a videorecording of their presentation. Working on grammar, vocabulary and spelling were also mentioned. We can see a lot of overlap in the answers to these two questions and it appears that several participants were already thinking as teachers when answering the question about their personal use of the videos for language learning purposes.

Discussion

This presentation of our preliminary results has focused on the perceived quality of AI-translated videos by native-speakers, their perceived impact on teachers-as-learners and their pedagogical potential as seen by a small cohort of language teachers. Despite the overall positive reactions, some reactions and comments raised awareness of possible difficulties using this type of technology for learning and teaching and drew attention to elements that may need to be taken into consideration before embarking on larger-scale pedagogical use of AI-translated videos. These will be discussed in this section.

A few of the participants remarked upon the strange, “otherness” effect or “cognitive dissonance” associated with seeing themselves speaking a language that they don’t actually understand. This should be both an invitation to further research and to precaution when working with students, for whom this type of strangeness could be psychologically disturbing. Many people do not like hearing their own recorded voice or seeing themselves in photographs or videos, even when all is “normal”. Thus, modified self-images could produce

heightened feelings of unease and require specific precautions to be taken. These might include just providing a forum to talk about the feeling evoked or go so far as to include referral to professionals for accompaniment. These have been taken into account in our Ethics application for the main phase of the study.

Two of our translated videos rendered the role of content and their incongruency with the linguistic output particularly interesting. In the first, the output was idiomatic to a local dialect spoken in Zagreb and its surroundings, which is neither a dominant way of speaking the language nor a standard. It could only be learnt either by living in that area or as a form of heritage language, when living abroad with family members. This however was in contradiction with the video content, as the participant stated having lived in various countries (France, Spain and England), none of them Croatia and considered herself to be Galician, but had never lived in Galicia. This caused the native Croatian speaker to remark:

The pace, the accent, the facial expressions, the way of speaking all fit a woman who was born or raised in Zagreb. It is impossible to tell that this is not a native Croatian speaker speaking in the video. Funny enough, this verisimilitude appears to be in contradiction with the narrative!

In another instance, the participant had spoken about growing up in Ekiti (a state in Southwestern Nigeria) and living in Lagos. Her recording had been translated into German, and the language reviewed by a native German speaker, who stated:

The second [recording] is very good as well, just the word choices are a bit awkward. [...] The content adds to that. If she had told us that she had never been to Ekiti, because she grew up in a German speaking country it might have passed, but then she talks about Lagos.

This type of reaction should no doubt alert us to coherence of discourse and verisimilitude.

While there are doubtless people who grew up in Lagos and who speak idiomatic German, they are likely to be few and far between. This type of example would probably be relevant only for very specific purposes.

By the same token, this should alert us to the necessity of discussing contextual appropriateness with our students. Imagine watching a pitch for wealth management with a background of a messy student bedroom, or a presentation on gastronomic food, situated in a carpark. Unless there were to be an explicit reason for creating these types of incongruencies, we might want to suggest that they be avoided. This also led us to re-evaluate the task we had set, for while talking about where one grew up may be a readily available subject that requires little if any preparation, it was obviously not appropriate for some of the target translations. For the next phase of our study, working with French students who will be translating their videos into English, we may want to take more time to work with them discussing the subjects they will present and thinking about cultural appropriateness.

Another potential difficulty, that one of the participants commented on above, is the inappropriateness or at least perceived uselessness of translated video for both true beginner language learners and for those whose are already perfectly fluent in the language being translated into. Indeed, it would seem that fluent speakers would prefer to record themselves directly in the Lx. However, regarding speakers with no or very little command of the language, it is our contention that those familiar with teaching beginners could imagine various possible activities using translated videos of self. As another participant indicated, her first step with a video of herself speaking an unfamiliar language was to try to detect words in this language that corresponded to things she had said in her own. We could also imagine using detection and reproduction of some very basic phrases that one could create for oneself as a means of introduction.

These suggestions serve also as a reminder of the plethora of possible pedagogical uses that our fifteen colleagues proposed. We can only imagine that any audience of teachers would come up with several more, depending on their own contexts of practice, experience and use of other technologies.

Our final remarks concern the tool itself. At the time when this experiment was conducted, some widely spoken languages, such as Yoruba, were not listed among the seventy-odd languages that *HeyGen* was able to translate from or into. This number has risen to 175 languages in the past year for the paid version. At present, we can also choose different accents for some languages such as English, Spanish or French. When this study began, this was only the case for English. Even if this was not an advertised possibility at the time, we discovered that *HeyGen*-translated videos sometimes contained regional specificities that were not listed as choices, but appeared in the translated videos. This was the case for French, Spanish and Croatian, where we could choose the language, but not the accent, yet found accented specificities in the translated videos (Québécois French, Peruvian or Mexican Spanish – our European native speaker was unable to tell – and Zagreb Croatian, as noted above). This leads us to question the models which were used to make these translations. Where do they come from? How were they chosen? Can we interpret them as acts of localisation or domestication, terms which are used in translation theory when something is adapted to a target language or location, accents being one of the ways to achieve that? Or, were these regional specificities, which were noticed by the native speakers in this study, just random occurrences linked to the availability of the models at hand? Although we cannot know for sure, it seems that the producers of *HeyGen* have reflected on the issue more thoroughly since our study began, as can be seen from this quote which is featured on their website at the moment of the writing of this article (November 2024): “175 + languages and dialects. The difference between Argentine Spanish and Mexican Spanish can make or break a video. That is why we cover almost all world languages and dialects.” The platform now also offers the possibility to edit everything that might seemed mispronounced or off: “Look through the original and translated versions and make precise line-by-line edits.” Our study compels us to acknowledge the rapidity in the evolution of the tool and raises questions about

methodology when one is doing research with such tools. If we were to translate these videos again using the exact same method, would we get the same results?

Conclusion

This preliminary study has enabled us to set out a theoretical framework for exploring the pedagogical affordances of AI-generated video translations. It has allowed us to run tests with a small group of participants in order to confirm the realism and accuracy of the translations thus produced and to explore with them some of the possible reactions students might have when confronted with this technology for language learning. Finally, it has allowed us to widen our own perspectives to envisage pedagogical uses we had not even imagined.

These results have brought us closer, but not provided any clear conclusions regarding our stated research questions, which we framed thus:

- Does seeing oneself speaking a foreign language of choice fluently induce changes in self-perception, in one's image of the "ideal L2 self", in motivation or in other affective states related to learning that language?
- Does taking oneself as a model influence the way in which a learner perceives and produces the foreign language?

They have, however, confirmed the validity of our inquiry and allowed insight into the protocols and methods that we will need to employ to address these questions with learners directly.

This is but the first step in a more extensive project using this technology that will be undertaken with language learners themselves. This next step will involve some 200 university students learning English and will delve further into the notion of the ideal L2 self and the specific ways in which video can enhance the visualisation of this self.

Bibliography

- Alm, A., & Watanabe, Y. 2023. Integrating ChatGPT in Language Education: A Freirean Perspective. *Iranian Journal of Language Teaching Research* 11(3), (Dec. 2023) 19-30.
- Cuq, J.-P., & Gruca, I. (2005). *Cours de didactique du français langue étrangère et seconde*. Presses Universitaires de Grenoble.
- de Meo, A., Vitale, M., Pettorino, M., Cutugno, F., & Origlia, A. (2013). Imitation/self-imitation in computer-assisted prosody training for Chinese learners of L2 Italian. *Proceedings of the 4th Pronunciation in Second Language Learning and Teaching Conference*, 90–100.
- Dörnyei, Z., (2009). The L2 Motivational Self System. In Dörnyei, Z., & Ushioda, E. (Eds.). *Motivation, Language Identity and the L2 Self*. Multilingual Matters, 9-42.
- Dörnyei, Z., & Ushioda, E. (Eds.). (2009). *Motivation, Language Identity and the L2 Self*. Multilingual Matters.
- Duff, P. (2000). Repetition in Foreign Language Classroom Interaction. In J. K. Hall & L. S. Verplaetse, *Second and Foreign Language Learning Through Classroom Interaction*. Routledge.
- Ellis, N. C. (2002). Frequency Effects in Language Processing: A Review with Implications for Theories of Implicit and Explicit Language Acquisition. *Studies in Second Language Acquisition*, 24(2), 143–188. <https://doi.org/10.1017/S0272263102002024>
- Ellis, R. (2012). *Language teaching research and language pedagogy*. John Wiley & Sons Inc.
- Henderson, A. J., & Skarnitzl, R. (2022). “A better me”: Using acoustically modified learner voices as models. *Language Learning & Technology*, 26(1), 1–21.
- Hilton, H. (2014). Mise au point terminologique: Pour en finir avec la dichotomie acquisition / apprentissage en didactique des langues. *Recherche et pratiques pédagogiques en langues de spécialité. Cahiers de l'Apliu*, Vol. XXXIII N° 2, 34–50.
- Huang, H.-C., & Eskey, D. E. (1999). The Effects of Closed-Captioned Television on the Listening Comprehension of Intermediate English as a Second Language (ESL) Students. *Journal of Educational Technology Systems*, 28(1), 75-96. <https://doi.org/10.2190/RG06-LYWB-216Y-R27G>
- Inceoglu, S. (2016). Effects of perceptual training on second language vowel perception and production. *Applied Psycholinguistics*, 37(5), 1175–1199. <https://doi.org/10.1017/S0142716415000533>
- Krashen, S. D. (1985). *The Input Hypothesis: Issues and Implications*. Longman.
- Léon-Henri, D. D. P. (2012). Teaching Foreign Languages through the Analysis of Film and Television Series: English for Legal Purposes. *Recherche et Pratiques Pédagogiques En Langues de Spécialité. Cahiers de l'Apliu*, Volume XXXI N° 2, 126–139. <https://doi.org/10.4000/apliut.2732>
- Rémon, J. (2012). Supports filmiques transversaux en pratique de la langue: Document authentique ou authentiquement intéressant ? *Recherche et pratiques pédagogiques en langues de spécialité. Cahiers de l'Apliu*, Volume XXXI N° 2, 90–102. <https://doi.org/10.4000/apliut.2702>
- Rodgers, M. P. (2013). *English language learning through viewing television: An investigation of comprehension, incidental vocabulary acquisition, lexical coverage, attitudes, and*

captions (Doctoral dissertation, Open Access Te Herenga Waka-Victoria University of Wellington).

- Ryan, S., & Irie, K. (2014). Imagined and Possible Selves: Stories We Tell Ourselves About Ourselves. In S. Mercer & M. Williams (Eds.), *Multiple Perspectives on the Self in SLA* (pp. 109–126). Multilingual Matters.
- Sweller, J., Ayres, P., & Kalyuga, S. (2011a). *Cognitive Load Theory*. Springer.
http://link.springer.com/chapter/10.1007/978-1-4419-8126-4_11
- Sweller, J., Ayres, P., & Kalyuga, S. (2011b). The Redundancy Effect. In *Cognitive Load Theory* (pp. 141–154). Springer New York. http://link.springer.com/chapter/10.1007/978-1-4419-8126-4_11
- Toffoli, G., & Allan, J. A. B. (1989). *Guided Imagery: Group Guidance for ESL Students*. LUGUS Productions Ltd.
- Vanderplank, R. (2016). *Captioned Media in Foreign Language Learning and Teaching: Subtitles for the Deaf and Hard-of-Hearing as Tools for Language Learning*. Palgrave MacMillan.
- Vanderplank, R. (1990). Paying attention to the words: Practical and theoretical problems in watching television programmes with uni-lingual (CEEFAX) sub-titles. *System*, 18(2), 221-234.
- Vanderplank, R. (1988). The value of teletext sub-titles in language learning. *ELT journal*, 42(4), 272-281.