Digital innovations in L2 motivation: Harnessing the power of the Ideal L2 Self

Svenja Adolphs[a,*], Leigh Clark[b], Zoltán Dörnyei[a], Tony Glover[c], Alastair Henry[d], Christine Muir[a], Enrique Sánchez-Lozano[c] and Michel Valstar[c]

Abstract

Sustained motivation is crucial to learning a second language (L2), and one way to support this can be through the mental visualisation of ideal L2 selves (Dörnyei & Kubanyiova, 2014). This paper reports on an exploratory study which investigated the possibility of using technology to create representations of language learners' ideal L2 selves digitally. Nine Chinese learners of L2 English were invited to three semi-structured interviews to discuss their ideal L2 selves and their future language goals, as well as their opinions on several different technological approaches to representing their ideal L2 selves. Three approaches were shown to participants: (a) 2D and 3D animations, (b) Facial Overlay, and (c) Facial Mask. Within these, several iterations were also included (e.g. with/without background or context). Results indicate that 3D animation currently offers the best approach in terms of realism and animation of facial features, and improvements to Facial Overlay could lead to beneficial results in the future. Approaches using the 2D animations and the Facial Mask approach appeared to have little future potential. The descriptive details of learners' ideal L2 selves also provide preliminary directions for the development of content that might be included in future technology-based interventions.

1. Introduction

Interaction with various forms of technology is increasingly becoming an everyday experience in the lives of people all over the world, and the rapidity of technological developments has created a constantly evolving landscape pushing the boundaries of what it is possible to achieve using technology. Videos – as well as images – can now be manipulated to make speakers appear to move and speak in ways that they do not when originally recorded, and this effect can even be achieved in real time (Thies et al., 2015). One

[*] Corresponding author. Telephone: 0115 951 5559
[a] School of English, University Park, University of Nottingham, NG7 2RD, UK
[b] School of Information and Communication Studies, University College Dublin, Dublin, Ireland
[c] School of Computer Science, Jubilee Campus, University of Nottingham, NG8 1BB, UK
[d] Department of Social and Behavioural Studies, University West, 461 86 Trollhättan, Sweden
Email addresses: svenja.adolphs@nottingham.ac.uk (S. Adolphs), leigh.clark@ucd.ie (L. Clark), zoltan.dornyei@nottingham.ac.uk (Z. Dörnyei), tony.glover@nottingham.ac.uk (T. Glover), alastair.henry@hv.se (A. Henry), christine.muir@nottingham.ac.uk (C. Muir), psxes1@nottingham.ac.uk (E. Sánchez-Lozano), michel.valstar@nottingham.ac.uk (M. Valstar).

positive practical possibility for language learners stems from the opportunity to digitally manipulate mental images, so that in a visual representation they appear as if their language proficiency had improved. Such developments could go hand-in-hand with various technological advances already available for use in language learning environments as a means of increasing the motivation of language learners (Stockwell, 2013). These may, for example, include virtual environments (VEs), such as Second Life and OpenSim (Gimeno-Sanz, 2016). In these virtual worlds, users can interact with their virtual selves or with avatars (e.g. Holzwarth et al., 2006). The potential power of such digitally generated realities is well reflected by the fact that successful experiments using avatars have involved improving people's financial saving through seeing a future version of themselves at a point where they may need to rely on their savings (Hershfield et al., 2011), or causing positive changes to healthy eating and exercise when viewing avatars that are engaging in health-conserving behaviours (Fox & Bailenson, 2009).

Thus, applications of digital technology can exert considerable motivational impact. The current paper pursues this agenda by presenting a possible new approach to motivating language learners through exploring the interface between digital animation technology and research on mental imagery and vision. Our underlying assumption is that a sustained level of motivation is fundamental to the successful mastery of a second language (L2) (Dörnyei & Ushioda, 2011), and the notion of vision has been proposed as a higher-order motivational factor that can help to fuel the language learning process in an ongoing manner (Dörnyei & Kubanyiova, 2014). As will be discussed below in more detail, one theoretical paradigm that can explain how imagery mediates motivation is the L2 Motivational Self System (Dörnyei, 2009), which relates vision to the language learner's future self-images. It has been suggested that if the person we would like to become is a competent speaker of a second language, an ideal L2 self can act as a powerful motivator because of the desire to reduce the discrepancy between one's actual and ideal selves (Higgins, 1987). The central idea in this paper is that with advancements in technology and avatar creation, it may be feasible to produce such possible selves digitally so that language learners can *see* and *hear* their virtual selves communicating successfully in the target language. That is, through taking advantage of technology, it can be possible to create images of the learner as an improved TL-speaker, and which can positively impact on their motivation much in the same way as an imagined ideal L2 self can do. This is, however, unchartered territory, and the current study offers a first step in exploring this intriguing avenue.

In experimenting with different methods of allowing L2 learners to physically see, rather than merely imagine, their ideal L2 selves, three different technology-based approaches were tried out. Approach 1 – 2D and 3D animations – used existing, off-the-shelf software to create an animated mock-up of the learners' faces. Approaches 2 and 3 – Facial Overlay and Facial Mask – created bespoke technological visualisations using videos of two short dialogues as a basis: in Approach 2, learners' facial appearance was overlaid onto one of the protagonists in the video dialogue ('blending' one to the other), and in Approach 3 learners' faces were 'floated' over of the target protagonist's, as though one would hold a mask over a face at a masquerade ball. Methods of creating these approaches are discussed, and are presented along with illustrations of how each was received when presented to participants. As a corollary, in the semi-structured interviews we conducted with participants, we also started to gather information about the social and career goals held by them, as well as the common speaker attributes that they reported aspiring to attain.

2. Literature Review

2.1 Motivating people through technology

Using technology to support language learning has become increasingly prevalent and has the potential to improve the motivation of language learners. As Stockwell (2013) highlights, this can work in two ways: a learner may either develop the motivation to interact with a technology and in turn be motivated to learn a language, or vice versa. Furthermore, technology use can be implemented by a teacher or used by learners outside of the classroom, through various means of computer-assisted language learning (CALL). Such initiatives have been promoted by the fact that in many industrialised countries, the use of computers and online resources has become commonplace in and outside of the classroom (Golonka et al., 2014). In this light, CALL can be broadly defined to encompass "any use of computer technology in the domain of language learning" (Hubbard, 2009, p. 2). This understanding widens the types of technology that CALL might include, such as interactive whiteboards, automatic speech recognition programs, virtual worlds, games, and the use of combined packages that utilise textbooks, CDs and online content. Arguably, combined packages like this have given way to a movement towards "atomized CALL" that sees individual technology being used for specific needs, such as Duolingo or virtual worlds such as Second Life and OpenSim (Gimeno-Sanz, 2016).

Technological developments have allowed for increased possibilities in the use of virtual environments (VEs) in CALL. In VEs, learners take control of digital avatars, and

either interact with the avatar themselves, or use it to interact with other users, the latter of which is increasingly being accomplished via voice chat (Hubbard, 2009). Thus, avatars can represent the virtual selves that users choose to interact with in VEs (e.g. Holzwarth et al., 2006), and as such, they can be designed for a range of purposes both in and outside of language learning. Examples include virtual stress counselling (Kavakli et al., 2016), non-human coaching (Warner, 2012), improving financial management (Hershfield et al., 2011), multilingual communication (Thorne, 2008), and language and literacy development (Thorne & Black, 2007). Well-designed avatars can help serve as a virtual "ideal social model" for a specific demographic of users, thereby acting as *motivational agents* (Baylor, 2011). Indeed, as Baylor argues, the most critical aspect of this process is the appearance of the virtual self, and it is of note that past research in this area has tended to focus on conceptualising and measuring a user's relationship with their avatar from a third person perspective (Behm-Morawitz, 2013).

Creating the right avatar appearance can be complex, and care needs to be taken to ensure people feel comfortable with using their avatars (Inkpen & Sedlins, 2011), as this may have serious bearings on the users' attitude and motivation (Hudson & Hurter, 2016). The degree to which people can identify with their avatar has a considerable impact on the motivational 'pull' of the technology. Previous research has indicated that people may identify more with avatars that represent their ideal selves, as opposed to those representing their real selves (Kim, Lee & Kang, 2012); moreover, users may establish a "wishful identification" with their avatar, in which they aspire to emulate attributes that their avatars possess (ibid.). This may hold potential in second language learning, where using imagined idealised visions of the learners' future selves has been recognised to be a powerful motivator (Dörnyei & Kubanyiova, 2014).

2.2 Visual stimuli and motivational arousal

The L2 Motivational Self System (Dörnyei, 2005; 2009) synthesizes various strands of past research in language learning motivation with advances in self-psychology, more specifically with the study of *possible selves* (Markus & Nurius 1986; for a review, see Dunkel & Kerpelman 2006) and the related theory of the *ideal self* (Higgins, 1987; 1998). The key aspect of the construct is the belief that a vivid image of oneself projected into a future situation where the L2 is successfully used serves as a motivational force. In the L2 literature this force has almost always been explained drawing on Higgins's (1987) self-discrepancy theory, arguing that a perceived discrepancy between the actual and the desired L2 selves

creates unease and an ensuing urge to reduce the gap by trying to approach the desired L2 self through motivated learning behaviour. Other scholars have conceived the motivational power of possible selves from a slightly different angle: Markus and Ruvolo (1989) considered desired future self-images to be *incentives*, Boyatzis and Akrivou (2006) saw the *hope* associated with a visualised future state as the main driver of action, while according to Hoyle and Sherrill (2006), possible selves "give rise to behavioural standards against which current self-representation is compared and with which it is reconciled through behaviour" (p. 1687).

Although there is no consensus about how exactly the visualisation of a desired future self-image exerts motivational influence, there appears to be a general agreement that ideal self-imagery *can* induce motivated behaviour. However, one of the reviewers of the current manuscript correctly noted that even if we accept this tenet to be the case, the concept of the ideal L2 self (a psychological construct) and a digital avatar (a visual/auditory stimulus) are two distinct entities, and therefore the learner's engagement with a visual scenario (i.e. with something one sees rather than imagines) may not automatically activate a future self-guide in the same way as mental imagery does, even if the avatar carries a close resemblance to the learner's appearance. Let us explore this issue further by examining the mechanisms by which a visual/auditory stimulus may evoke a motivational impulse.

In trying to explain the motivational power of mental imagery, Dörnyei and Kubanyiova (2014) have reviewed research in sport psychology on the behaviour-boosting influence of visualisation. Several performance-promoting techniques have been developed in this domain under the neurobiological premise that, in a broad sense, the brain cannot distinguish between an actual physical event and a vividly imagined scenario of the same event. One frequently employed method of energising long-term training efforts has been to get athletes to visualise their desired goals; for example, many professional athletes would regularly envisage winning a race or stepping on the top of the podium as a motivational exercise (e.g. Cumming & Ste-Marie, 2001; for a review, see Gregg & Hall, 2006).

Even more relevant to the current study is the use of video tapes to develop constructive imagery skills. In a volume on sport psychology published by the American Psychological Association, Gould, Damarjian and Greenleaf (2002) explain that "success tapes", which contain some of the athletes' best competition moments or some edited, perfect performances, can have an arousing effect: "For example, the 1998 U.S. women's ice hockey team used an imagery tape of season highlights interspersed with Muhammad Ali clips to create positive feelings of confidence before their gold medal-winning game" (p. 65). As the

researchers continue, the athletes' favourite music can be "dubbed onto the tape to serve as a cue to trigger excellent performance in the future" (ibid), and listening to this music on a portable stereo headset can be helpful to trigger feelings of success in situations in which it is not practical to watch a videotape (e.g., at a competitive venue) (ibid).

Motivated by the high stakes and the potential of considerable financial gains, sport psychology has produced a solid body of research on the applicability of various forms of visual stimuli for enhancing performance. This being the case, the question is how transferrable the psychological context of sports performance – a domain determined to a large extent by physical and motor skills – might be to the more cognitively determined area of SLA. In this respect, Allan Paivio's classic theory of imagery functions is relevant. In a seminal paper analysing sport contexts, Paivio (1985) argued that imagery has both cognitive and motivational functions, with the former concerning the rehearsal of specific sport skills, and the latter involving imaging oneself coping effectively in challenging situations or focusing on various arousal experiences in order to gain energy and confidence. Thus, some of the psychological functions associated with sport imagery that Paivio identified went beyond the promotion of sport skills alone and offered viable opportunities of transferability to other domains. Just to take one particularly relevant illustration of this, the effective use of video tapes has also been validated in psychotherapy, where, according to Holmes and Matthews (2010), an established strategy for treating social phobia is to require patients "to contrast their mental image of themselves with a video of their own social performance, while being encouraged to view it as if watching another person" (p. 358). As they argue, the critical mechanism involves making an alternative and more positive image "more accessible, and thus more effectively [able to] compete with the original maladaptive image(s)" (ibid).

We need to reiterate that the functional link between visual and mental imagery is related to the fact that the same brain regions that encode incoming sensory information are involved in mental imagery (see e.g. Cattaneo & Silvanto, 2015; Pearson & Kosslyn, 2015). A particularly clear example of the parallel functions concerns emotions. In their influential paper on harnessing mental simulation, Taylor et al. (1998) submitted that "A major consequence of mental simulation is the evocation of emotional states and their potential control. Imagining a scenario does not produce a dry cognitive representation but rather evokes emotions, often strong ones" (p. 431). Indeed, imagining a scenario can evoke an emotional state that is as powerful as physically experiencing the same scenario, and for example the core characteristics of post-traumatic stress disorder (PTSD) are the powerful emotions that are produced by imagery in the form of 'flashbacks' to the original traumatic

event (Holmes & Matthews, 2010, p. 350). Based on this correspondence of emotional impulses in real and imaged situations, carefully crafted visual stimuli can be utilised to create positive emotional responses to L2 learning situations, thereby increasing the learner's psychological willingness to expose themselves to such stimuli. In this way, avoidance tendencies can in effect be reversed and transformed into approach tendencies.

Finally, it needs to be emphasised that the existence of an ideal L2 self in and of itself is not sufficient to ensure that motivated action will follow. Multiple conditions have been highlighted that may be required for this to take place (for a review, see Dörnyei & Ushioda, 2011), among them that a learner must have an ideal L2 self-image, and that it must be accessible and regularly activated within an individual's working self-concept (Markus & Nurius, 1986; Markus & Kunda, 1986; Markus & Wurf, 1987; Oyserman & Destin, 2010). A possible self must also be plausible and perceived as within reach (Oettingen & Thorpe, 2006), and sufficiently different from an individual's current self (Higgins, 1987). Of importance to the current study, it has also been demonstrated that the level of abstraction of an ideal L2 self is also important. It is generally agreed that specific images must be vivid and well elaborated (e.g. Dörnyei & Kubanyiova, 2014), yet mental images lacking clarity and definition have also been found to be motivational if, for example, they represent an abstract yet still highly desired capacity (Trope & Liberman, 2010). Indeed, Eitam and Higgins (2010) have argued that the degree to which a mental representation is activated to guide processes of thought and behaviour is dependent on the representation's "motivational relevance" (p. 951). A highly realistic representation of being *engaged* in the desired target activity – such as seeing oneself successfully interacting in the L2 – is likely to have a high degree of motivational relevance, and therefore substantial motivational power.

2.3 Designing virtual selves and research objectives

The design of virtual selves and avatars can range from 2D models created from as little as one photograph, through to advanced processes of motion capture for animations (Szczuko et al., 2009). A number of off-the-shelf programs exist that can be used to create models for use as avatars. A study attempting to design a virtual stress counsellor system provided a number of examples, including Xist Face Tracking and Motion Builder, Mascard and 3D Max, and Crazytalk (Kavakli et al., 2016). Some programs offer more lifelike models than others, which can be preferable to more cartoonish alternatives that users may perceive as being non-credible representations (Warner, 2012).

The initial process of creating a digital 'self' can be relatively quick. Software such as CrazyTalk (Reallusion, 2016) can create models with relative ease using a limited set of reference points from a target's face (Verdet & Hennebert, 2008). Other programs, such as 123D Catch (Autodesk, 2016), allow users to create lifelike 3D models of virtually any object, using a computer or smartphone application (Chandler & Fryer, 2013).

While the potential motivational impacts of avatars and VEs have shown promise, there is less research on combining virtual and real-life contexts, and within the context of ideal L2 self research this has likewise not yet been addressed. In creating initial models of virtual selves, the current study uses existing digital animation software (Approach 1) and bespoke facial 'replacement' techniques (Approaches 2 and 3) with the aim of highlighting possible opportunities and challenges for the uses of these technological developments in the context of language learning. Our main objective has been to understand the viability of digital visualisation for the purpose of promoting L2 motivation, to explore study participants' ideal L2 selves, and to gather information concerning the choice of context of appropriate L2 speaking scenarios.

3. Method

3.1 Participants and design

Participants were seven female and two male Chinese postgraduate students currently engaged in further study in the UK. The current study chose to focus on Chinese L1 speakers because they represent the largest single group of English language learners in the world – over 390 million in Mainland China alone (Wei & Su, 2012) – and form one of the biggest markets for dedicated language learning pedagogy and materials. The use of CALL is also seen as being particularly beneficial in Chinese EFL classrooms (Li & Walsh, 2011). Furthermore, the L2 motivational self system and the ideal L2 self have previously been validated in this context (e.g. You & Dörnyei, 2016; You, Dörnyei & Csizér, 2016).

Data was collected over a three-stage interview process, conducted with each participant individually (with the exception of one participant who only took part in the first two interviews). In total, 26 interviews were conducted, each lasting about one hour on average. The interviews from each stage were subsequently transcribed. Thematic analysis was used as a means of identifying emerging themes (Nowell et al., 2017).

3.2. The interviews

The first round of interviews aimed at understanding the learners' ideal L2 selves and their language goals. The purpose was to look for common scenarios in which learners' visualised

their ideal L2 selves. Participants completed a 'language goals tree' (adapted from Hock et al., 2006 and Magid & Chan, 2012) to encourage them to think in greater detail about more specific language goals. The tree was divided into professional/academic goals and social/leisure goals. The goals listed by the participants were organised into common categories (see Section 4.4).

In the second round of interviews, participants were shown a series of curated video clips based on the language goals they had described in their first interview. For example, if participants described giving an academic presentation in English as a goal, then they were shown video clips of people giving academic presentations. The purpose was to investigate whether, and in what ways, these videos were representative of the ideal L2 selves the participants had previously described. Specifically, participants were encouraged to describe what elements of each video they felt were particularly important to them. The selected videos featured a variety of different L1 English speakers, as well as several L2 English speakers, including JK Rowling providing a university commencement speech[1] and a Chinese student discussing their experience of a university in the UK[2]. In total, a pool of 54 separate video clips was compiled from analysis of the first interviews, and between 20-30 relevant videos, each approximately 30 seconds in length, were shown to each participant. The order of videos was randomised for each participant, and an average of 24 were viewed in each session (the number depending on the amount of discussion each led to).

In the third round of interviews, participants were shown digital representations of their ideal L2 selves. Here, discussion centred on assessing the feasibility and the perceived benefits and drawbacks of each version in a context of 'seeing themselves' speaking fluently in English. As already mentioned, the digital representations involved three design approaches, details of which follow below.

3.2.1 Approach 1: 2D and 3D digital animation process

Approach 1 utilised the off-the-shelf software package CrazyTalk 8 (Reallusion, 2016) to design animated 2D and 3D models of the participants. Participants viewed these animations as they were, i.e., these animations were not seen within the context of real-life interaction or integrated with other video data. Two images of each participant were imported into the programme, facial photographs taken from both the front and the side (for the 2D version only the front facing photograph was used). For both 2D and 3D models, specific facial

[1] https://www.youtube.com/watch?v=wHGqp8lz36c&t=50s
[2] https://www.youtube.com/watch?v=G1Cd96bcw-s

points (e.g. the corners of lips and eyes) were manually identified. For the 3D models, the computer generated avatars were then processed manually to ensure that the output reflected participants' likenesses as far as was possible. For example, this included changing the shape of the face and the skin tone (see Figures 1 and 2 for examples).

[INSERT FIGURES 1 AND 2 HERE]

Ultimately, Approach 1 resulted in three variations: one 2D model, and two 3D models, one with and one without context. The sole 2D model remained similar in appearance to the initial photograph; the mouth and eyes were animated, and the model simulated some head movements. The 3D models were both fully computer-generated and placed the head upon the upper body of a digital model. The base output of the 3D model lacked the addition of contextual features such as hair, glasses, clothing and background. The second 3D model included these features. However, while they were matched to create as close a representation as possible, this was limited by the fixed selection of options available in the software.

Once all three examples for Approach 1 had been created, audio files were imported so that when these animations 'spoke', it gave the illusion that they were proficient English speakers. The texts chosen for this were of casual conversations between two people in authentic contexts (while waiting to watch a band play a concert), and were taken from an English language-learning textbook (McCarthy, McCarten & Sandiford, 2014) prized for its reliance on research from the Cambridge English Corpus (2017). The dialogue included both male and female audio, and was matched with the gender of the participant.

3.2.2 Approach 2: Facial Overlay process

Approach 2 overlaid the image of the participant onto the video scenario, thereby 'blending' together the appearances of the learner and target protagonist. Four videos were filmed by the research team, each featuring a dialogue between two L1 English speakers, one male and one female. The dialogue for these scenarios was the same as described above (the audio from these videos was extracted and used for Approach 1). The videos were approximately 20 seconds in length. Two were filmed against a white wall (referred to as 'without context' for their lack of noticeable recognisable surroundings) while another two were filmed at an outdoor picnic table (referred to as 'with context'). Learners were shown both context types. Either the male or female protagonist could be chosen as the target onto which the participants' facial features were overlaid.

The process of tracking for the Facial Overlay started with detecting a key set of facial points in the photographs (see Figure 3 for the boundaries of these points), and then by identifying these same points in the video scenarios. Using these key points, a set of correspondences, or mappings, between each video frame and the target image were given, and were used to transfer the facial appearance from the image to the video (for an example see Figure 4). The use of a set of 66 points allowed the appearance transfer to be locally accurate, and robust to changes in pose and expression. Using this process, the participant's face was therefore 'overlaid' onto the selected protagonist (e.g. Li et al., 2016; informed also by Cosker, Krumhuber & Hilton, 2011).

[INSERT FIGURES 3 AND 4 HERE]

There were some limitations in this process. In the configuration used, this approach did not adequately handle differences in illumination and lighting conditions, and therefore the videos had to be created in grayscale. Furthermore, given that the transfer was done through an inner triangulation, the information outside of the outer points (e.g. hair, neck) was left out of the transfer process, leading to what was in places an abrupt change of appearance. Finally, there were some noticeable visual artefacts that arose when faces were transferred onto the male protagonist in the videos due to his wearing glasses.

3.2.3 Approach 3: Facial Mask process

Approaches 2 and 3 both involved using facial tracking (e.g. Sánchez-Lozano et al., 2016), with the aim of transferring the facial likeness of participants from their photographs onto a target in a pre-filmed video scenario. In Approach 3 the protagonists' faces in the videos were instead replaced with a 'cut-out' of participants' images, thus creating a mask-like effect (for an example see Figure 5). While this 'mask' moved with the protagonists' faces as they moved throughout the dialogue in the videos, it lacked any further animated capabilities. This process used a small subset of 6 facial points to deduce all facial movements. Points at either side of the head were used to work out the sideways tilt of the face, and points between the forehead and nose were used to work out the forwards and backwards tilt. These points were also used to derive the specific positions the masks should be placed within in each video. The difficulties in illumination and lighting were not encountered in this process, and therefore video outputs were created in colour.

[INSERT FIGURE 5 HERE]

4. Results

4.1 Evaluation of the three technologies – Approach 1: 2D and 3D digital animation feedback

In evaluating the three technologies that were used in Approach 1, there were notable differences in feedback towards the 2D and 3D animations. For the 2D images (in which a photograph was digitally animated to move the head, eyes, and mouth only) the animated embellishments appeared to actually *reduce* how real and potentially motivating the images were perceived to be. This was predominantly attributed to the movements of the eyes and mouth, particularly when the image was 'speaking' through the imported sound files:

> *It looks a bit weird… because of the eyeballs.* (P4)
>
> *The movement of the mouth…it looks like some kind of stranger.* (P6)
>
> *The eyes and mouth look like the real person isn't speaking.* (P7)

The 3D computer generated images were perceived as being more natural and realistic, despite the 2D images being created from a real photograph of the learners' likeness. Again, the movement of the image was reported of particular importance, and in the 3D images the mouth, eyes and head movement were viewed as less dissonant from the rest of the image in terms of its naturalness, which reflected positively in the 3D images overall:

> *It's more natural I think, the movement of the head and other things*. (P5)
>
> *The eyes and mouth are better than the first one...it's more natural*. (P7)
>
> *I think the 3D is definitely more real than 2D to start with.* (P8)

Perhaps unsurprisingly, the 3D models augmented from the initial base output (to include e.g. hair, clothing and background images) were preferred to the base output models. Learners found these images more relatable than the other two alternatives presented in Approach 1:

> *She looks like me.* (P8)
>
> *This is much better, the hair and the background.* (P9)

4.2 Approach 2: Facial Overlay feedback

The videos included in Approaches 2 and 3 were filmed both indoors ('without context') and outdoors ('with context'), and participants' responses to this aspect in both Approaches 2 and 3 were more positive about the latter. As one participant describes, this environment was more like 'our daily lives':

> *The scenario is more natural.* (P3)

The situation looks more like our daily lives…the background is so familiar. (P7)

More likely to put yourself in that context. (P9)

Despite the inclusion of a more familiar environment, the effectiveness of this approach was diminished by the sometimes inconsistent blending of the learner and the protagonist's appearances in the video scenarios. Some features of both were visible in the final image as the facial overlay included a limited area of the target's face. Participants reported both that it would be more 'comfortable' to watch the original video rather than one augmented with the Facial Overlay, and highlighted the difficulty in relating to the image as a representation of themselves owing to the confused question of identity this created. However, these issues may be alleviated by improving the quality and realism of the image in future developments, and participants were open to engaging with future iterations of this approach:

Even though I don't recognise this person, if it's made more real I can relate to them. (P8)

Similarly negative feedback was received regarding the naturalness of the image. This was highlighted when viewing the 2D digital animation from Approach 1, and again criticism was particularly related the eyes and mouth:

It's a bit weird… I think the mouth. (P3)

I prefer the last one [3D animation]. It looks more natural. (P7)

When comparing the Facial Overlay process of Approach 2 to the variations of Approach 1, the 3D digital animation continued to be perceived most favourably. Participants suggested that fully computer-generated images may prove to be more motivational, at least until the technology of the Facial Overlay process might be improved. It was also suggested that a combination of a virtual self in a real-life video context may combine the best aspects of these two approaches:

I'd rather have a virtual me – at least that's based on me rather than based on somebody else… it's not exactly the same but I know this virtual me is based on myself… so long as it's a real context. (P4)

Participants suggested that having a virtual self modelled on their own appearance may be more relatable to than the 'blending' of two people's appearances as with Approach 2's Facial Overlay process, perhaps in turn also mitigating potential 'uncanny valley'-like effects.

4.3 Approach 3: Facial Mask feedback

While the inclusion of colour in the Facial Mask process was appreciated, feedback on this approach was often immediately dismissive:

> *Like wearing a mask…not natural because the image of the head is covered.* (P3)
>
> *Distracting, because I was paying attention to the movement.* (P5)
>
> *You don't have facial expression.* (P9)

The main positives of this approach were directed towards the video being in colour, as opposed to the grayscale necessitated in the Facial Overlay process of Approach 2 ("*Background is really important… this [with context] is better because it's colourful*" *(P2)* / "*I like the idea of the natural context*" *(P4)*). However, the size and movement of the Facial Mask was distracting, and provided no animated capabilities beyond it being tracked onto the movement of the protagonist's face. As such, there were no suggestions by participants that this approach might have any motivational value were it to be developed further.

4.4. Imagined contexts and scenarios

In analysing the completed 'language goals trees', a varied range of desired goals were described by participants. These were organised into sixteen categories, further divided to separate academic and career focused goals, social and leisure oriented goals, and goals that might be applied to both (see Tables 1 and 2). Being postgraduate students, understanding more of the language used in their course material, particularly lectures and literature, was a common goal that participants wished to achieve in their future selves:

> *Sometimes it's very hard for me to understand, especially when they use some jargon.* (P1)
>
> *There are some professional words I can't understand, when I check words in some books in the library I can't get the meaning.* (P2)

Table 1: Examples of one participant's language goals

Academic & Career	Social & Leisure
• Master terminology in my study field • Write high quality papers • Improve presentation skills • Read English textbooks more easily	• Enjoy movies & music without subtitles or lyrics • Understand & tell jokes • Talk freely with friends

• Have no difficulty understanding lectures	• Have no difficulty communicating in everyday life (e.g. banks & restaurants) • Make phone calls

Table 2: Categories of participants' language goals for their ideal L2 selves.

Academic & Career	Both	Social & Leisure
• Understand lectures and speeches • Talk with institutional staff & superiors (boss, supervisor) • Give a speech or presentation • Improve writing ability • Find a job	• Understand nuanced word choice and prosody • Speak fluently and naturally • Improved reading ability • Converse with peers and friends • Travel and work abroad	• Engage with local people in the UK • Make friends and discuss their cultures • Conduct daily life tasks (e.g. shops, bank) • Understand and employ humour • Watch films and other media without subtitles • Order food in a restaurant

Conversing with their peers and classmates, as well as their friends outside of their studies, was also commonly reported. Outside of university life, participants wished to engage with local people in the city and the country in which they were living, including when conducting daily business such as going into shops and ordering food in restaurants. Participants indicated that limited linguistic ability in these common situations, such as being in restaurants, can result in social tension, frustration or embarrassment:

> *I am afraid in the restaurant because I look (at) the menu and what a mess. I don't know what they're talking about, I don't know what the dish is, I don't know the taste… I imagine speaking fluently but when I open my mouth, it frustrates me.* (P1)

> *When I order some meals, I can't understand. I think they [restaurant staff] should be angry with me. I just spend lots of time (saying) pardon, sorry. It makes me feel embarrassed.* (P2)

Understanding and using humour was also commonly discussed by participants as a socially-driven desirable trait, as lacking humour may isolate them in social situations like watching a film or interacting with their peers in class:

> *I understand most of it, but they (characters) say a joke and everyone laughs except me! I don't know why they're laughing.* (P1)
>
> *When classmates tell some jokes, I can't understand why they're laughing. It's embarrassing, actually.* (P9)

The observed patterns of ideal L2 selves suggest that while some goals are particular to an individual (e.g. communicating at the gym as a part of their daily life tasks), many goals emerged that were common to these English language learners. In developing appropriate contexts for further technology-based visualisations of ideal L2 selves, identifying these common goals may help in focusing on a more manageable range of scenarios that are still relevant for a wide number of language learners.

4.5. Desired spoken features and attributes

Common themes emerging from learners' ideal L2 selves were also observed when participants discussed them in the context of the videos they viewed of both L1 and L2 English speakers. Participants often referred to both the vocal qualities of the speakers in the videos, and the manner in which they spoke. These observations centred on the following concepts: accent, confidence, fluency, naturalness, speech rate and use of gestures. Some participants displayed aspirations towards achieving accents similar to the L1 English speakers in the videos, suggesting that for some participants the 'near-peer role models' (i.e. advanced Chinese English speakers) acted as less effective role models; however, achieving a native-like accent was not perceived as either realistic or achievable:

> *I see good English speaking, but I don't like (the) Chinese accent.* (P6)
>
> *I think that maybe it's very hard for me to have the native accent, so I just want to be fluent and accurate.* (P9)

The attributes highlighted were not limited only to the language a speaker used, but also to their personality and how they presented themselves within the videos (e.g. being comfortable on stage when giving a presentation). Having an appropriate speech rate was frequently commented on, and this was also related to fluency and naturalness. The videos that were spoken at a slower rate (having been designed specifically for language learning audiences) were not perceived as related to the participants' ideal L2 selves, even though participants were more easily able to understand them:

In real situations, the speed of your speech is not so slow. (P5)

I can understand what they say when they speak slowly… I want to speak quicker like a native speaker. (P6)

Regarding the use of gestures, this was discussed in relation to 19 of the 54 videos, 14 of which were presentations or speeches. Participants discussed both what they perceived to be appropriate and inappropriate uses of gesture in the videos. While discussion of gesture primarily focused on the speakers' hands, further inquiry into this specific aspect of learners' ideal L2 selves may create a more thorough understanding of these issues. This type of multimodal approach would be valuable when pursued alongside the development of this technology:

I like to be like her…with some hand gestures to express me (sic) to the audience. (P2)

Ideal L2 selves were also discussed in terms of the communicative contexts that the speakers appeared in, and in this case were strongly linked to the language goals participants' described in the first interviews. It was often the case that certain details did not match participants' imagined ideal L2 selves, for example participants who described giving speeches commented on the size of the audience being too large or too small:

The audience size is too little. (P2)

He has a large audience but I never thought about that before. (P9)

Although many questions remain unanswered in this respect, findings suggested a level of commonality between participants regarding those language and communication features they believed to be most relevant to their ideal L2 selves. This offers an initial positive indication of the feasibility of curating a common set of videos for learners for use with such technology, and although it is important that learners be allowed to choose the videos they most identify with, it might be possible to create this from a common 'pool' of content.

5. Discussion

This paper has introduced a novel approach that involves visualising a language learner's ideal L2 self through the use of technology. The observations put forward by participants revealed a number of design challenges in finding the balance between the realism and relatability of the images, and in relation to the context in which the image is situated.

5.1 Visualising ideal L2 selves

Overall, participants displayed a preference for the 3D animated images from Approach 1 that were personalised to their own appearance, even though this was done with generic templates that were sometimes quite far from the participants' actual likeness. Despite being used with real-life video data and, in one of the videos, a relatable background environment, Approach 2's Facial Overlay process appeared too unrealistic in its current state for the learners to be able to identify with. Deciding on how much of the learner and video target to facially map in this way, and improving the quality of movement in features such as the mouth and eyes, are all aspects of this approach that can be improved upon.

One aspect of the technology used in Approach 2 that is particularly challenging is matching the image of the learner with that of the protagonist in the video: issues related to facial structure, gender and skin tone all need to be appropriately resolved in order for the image to be both relatable and realistic to learners, and thus to provide any motivational potential. It may be, however, that such developments trigger negative, 'uncanny-valley'-like effects, and the configuration of these features demands further investigation.

The digitally-embellished 2D images in the use of Facial Masks in Approaches 1 and 3 both appear to hold little value for further investigation. Further exploration of the use of 3D virtual avatars, however, may prove to be a useful endeavour, possibly to be presented in combination with real-life contexts. Investigating different combinations of virtual and real selves, as well as of backgrounds and contexts, will be crucial for refining these technological approaches to 'visualising' an ideal L2 self for the purposes of increasing language learner motivation.

5.2 Developing relevant L2 scenarios

With regard to creating appropriate content for the ideal L2 self scenarios, findings indicated that there were a number of observable patterns common to the ideal L2 selves of participants in the sample, and which related to goals in both academic and social lives. While there was a certain amount of diversity amongst the participants – for example regarding which shops would form part of a learner's daily routines, or the type of presentation they might give and to whom – it was possible to categorise these effectively into broader rubrics. This suggests that it may be a realistic task to curate or film a selection of videos that will be of relevance to a considerable number of language learners, thereby forming an effective base which learners will be able to personalise when creating their own L2 avatars. Exploring datasets from larger

and more varied participant groupings is an important next step in refining and adding to the categories and desired learner attributes presented in this paper.

The interviews demonstrated our participants' abilities to critique the curated videos with respect to the extent to which they aligned with their ideal L2 self visualisations. However, the videos were admittedly somewhat limited in their diversity; while some of the videos featured English speakers, and with some of these being of the same nationality as the participants, a greater variety of target protagonists will be needed in future studies. Participants were able to discuss accents and gestures, as well as the contexts in which the videos were situated. However, some of the emerging ideas were more abstract or general than others, thus requiring further research to achieve practical solutions. Further analysis of gestures in particular could provide greater insight into multimodal aspects of learners' ideal L2 selves.

In sum, the examined issues – along with others which we did not examine in this initial exploratory study – have important ramifications for utilising digital technology in helping learners to create ideal L2 selves. From the participants' evaluations it is clear that it is often seemingly secondary details that determine the extent to which learners will be inclined to engage with these innovative technologies for maximum motivational gain. Finally, it should also be remembered that similar to how a learner's ideal L2 self will change and develop over time, so too must the personal avatars used in the technologies.

5.3 Further limitations

In addition to the limitations highlighted above, we would like to acknowledge a number of further limitations. The sample size for this study was small, coming from a single nationality group; even though we have argued for the likely applicability of these findings to other learner populations, this of course requires empirical support. Two of the learners in this study did not consent for their pictures to be taken, and so in the third round of interviews these participants viewed and discussed images that were not of themselves. Due to the fact that a central theme of this interview was to discuss the degree of relatability of each approach to the learner, it is likely that this has impacted on their responses (although their responses were informative in the context of the exploratory aims of the current study). Finally, although our research team included IT specialists, it became clear during the investigation that the level of IT involvement which would have been required to produce technological solutions that would have eliminated all the emerging participant reservations –

some predictable, but some totally unexpected – went beyond the anticipation of the initial research design (and budget!). This may also explain to some extent why digital innovations tend to be slow in the domain of SLA, since setting up the necessary mixed-expertise researcher groups poses considerable motivational and logistical challenges.

6. Conclusion

This paper investigated how technology-based approaches might be used to create digital versions of learners' ideal L2 selves, and how this approach can prime effective visualisation processes for L2 learners. Observations have highlighted two approaches as having particular future potential; 3D animation (from Approach 1) and the Facial Overlay process (Approach 2). Animated 2D models and Facial Masks were received less favourably. Initial steps were also taken to provide a description of the learners' specific ideal L2 self-images, as well as how it may be possible to develop appropriate content and contexts for this type of application. We would like to stress that our study represents only the beginning of a journey to understand the potential use of technology to enhance or scaffold L2 motivation by creating a digitally augmented image of the ideal L2 self. Besides offering results concerning practical implications and implementations, our study lays the foundation for two primary issues that future research is likely to further address: the general motivational value of different technological innovations and approaches, and, in particular, the relationship of digitally enhanced images to mental visualisations of possible self-images. As mentioned earlier, while these are largely uncharted areas, it is clear that these lines of inquiry will assume increased importance in the changing educational and technological landscape of language learning in the twenty-first century.

References

Autodesk. (2016). *123DCatch* [Computer software]. California, USA.

Baylor, A. (2011). The design of motivational agents and avatars. *Educational Technology Research and Development*, 59(2), 291-300. DOI: 10.1007/s11423-011-9196-3

Behm-Morawitz, E. (2013). Mirrored selves: The influence of self-presence in a virtual world on health, appearance, and well-being. *Computers in Human Behavior*, 29(1), 119-128. DOI: 10.1016/j.chb.2012.07.023

Boyatzis, R.E., & Akrivou, K. (2006). The ideal self as the driver of intentional change. *Journal of Management Development, 25*(7), 624-642.

Cambridge English Corpus. (2017) retrieved from http://www.cambridge.org/gb/cambridgeenglish/better-learning/deeper-insights/linguistics-pedagogy/cambridge-english-corpus

Cattaneo, Z., & Silvanto, J. (2015). Mental imagery, psychology of. In J. Wright (Ed.), *International encyclopedia of the social and behavioral sciences* (2nd ed., Vol. 15, pp. 220-227). Oxford: Elsevier.

Chandler, J., & Fryer, J. (2013). Autodesk 123D catch: how accurate is it. *Geomatics World*, *2*(21), 28-30.

Cosker, D., Krumhuber, E., & Hilton, A. (2011). A FACS valid 3D dynamic action unit database with applications to 3D dynamic morphable facial modeling. In *Computer Vision (ICCV), 2011 IEEE International Conference* (pp. 2296-2303). IEEE.

Cumming, J.L., & Ste-Marie, D.M. (2001). The cognitive and motivational effects of imagery training: A matter of perspective. *The Sport Psychologist, 15*(3), 276-288.

Dörnyei, Z. (2005). *The psychology of the language learner: Individual differences in second language acquisition*. Mahwah, NJ: Lawrence Erlbaum.

Dörnyei, Z. (2009). The L2 motivational self system. In Z. Dörnyei & E. Ushioda (Eds.), *Motivation, language identity and the L2 self* (pp. 9-42). Bristol: Multilingual Matters.

Dörnyei, Z. & Ushioda, E. (2011). *Teaching and researching motivation* (2nd ed.). Harlow: Longman.

Dörnyei. Z., & Kubanyiova, M. (2014). *Motivating students, motivating teachers: Building vision in the language classroom*. Cambridge: Cambridge University Press.

Dunkel, C., & Kerpelman, J. (Eds.). (2006). *Possible selves: Theory, research, and applications*. New York: Nova Science.

Eitam, B., & Higgins, E.T. (2010). Motivation in mental accessibility: Relevance of a representation (ROAR) as a new framework. *Social and Personality Psychology Compass 4*(10), 951-967.

Fox, J. & Bailenson, J.N. (2009). Virtual self-modeling: The effects of vicarious reinforcement and identification on exercise behaviors. *Media Psychology, 12*(1), 1-25.

Gimeno-Sanz, A. (2016). Moving a step further from “integrative CALL”. What's to come?. *Computer Assisted Language Learning, 29*(6), 1102-1115.

Golonka, E.M., Bowles, A.R., Frank, V.M., Richardson, D.L., & Freynik, S., (2014). Technologies for foreign language learning: a review of technology types and their effectiveness. *Computer Assisted Language Learning, 27*(1), 70-105.

Gould, D., Damarjian, N., & Greenleaf, C. (2002). Imagery training for peak performance. In J.L. Van Raalte & B.W. Brewer (Eds.), *Exploring sport and exercise psychology* (2nd ed., pp. 49-74). Washington, DC: American Psychological Association.

Gregg, M., & Hall, C. (2006). Measurement of motivational imagery abilities in sport. *Journal of Sports Sciences, 24*(9), 961-971.

Hershfield, H.E., Goldstein, D.G., Sharpe, W.F., Fox, J., Yeykelis, L., Carstensen, L.L., & Bailenson, J.N. (2011). Increasing saving behavior through age-progressed renderings of the future self. *Journal of Marketing Research, 48*(SPL), S23-S37.

Higgins, E.T. (1987). Self-discrepancy: A theory relating self and affect. *Psychological Review, 94,* 319-340.

Higgins, E.T. (1998). Promotion and prevention: Regulatory focus as a motivational principle. *Advances in Experimental Social Psychology, 30,* 1-46.

Hock, M.F., Deshler, D.D., & Schumaker, J.B. (2006). Enhancing student motivation through the pursuit of possible selves. In C. Dunkel & J. Kerpelman (Eds.), *Possible selves: Theory, research and application* (pp. 205-221). New York, NY: Nova Science.

Holmes, E.A., & Matthews, A. (2010). Mental imagery in emotion and emotional disorders. *Clinical Psychology Review, 30,* 349–362.

Holzwarth, M., Janiszewski, C., & Neumann, M.M., (2006). The influence of avatars on online consumer shopping behavior. *Journal of marketing, 70*(4), 19-36.

Hoyle, R.H., & Sherrill, M.R. (2006). Future orientation in the self-system: Possible selves, self-regulation, and behavior. *Journal of Personality, 74*(6), 1673-1696.

Hubbard, P. (Ed.) (2009) *Computer Assisted Language Learning: Critical Concepts in Linguistics*. Routledge.

Hudson, I., & Hurter, J. (2016, July). Avatar Types Matter: Review of Avatar Literature for Performance Purposes. In *International Conference on Virtual, Augmented and Mixed Reality* (pp. 14-21). Springer International Publishing.

Inkpen, K.M., & Sedlins, M. (2011, March). Me and my avatar: exploring users' comfort with avatars for workplace communication. In *Proceedings of the ACM 2011 conference on Computer supported cooperative work* (pp. 383-386). ACM.

Kavakli, M., Ranjbartabar, H., Maddah, A., & Ranjbartabar, K. (2016). Tools for eMental-Health: A Coping Processor for Adaptive. *Integrating Technology in Positive Psychology Practice*, 127-162.

Kim, C., Lee, S.G., & Kang, M. (2012). I became an attractive person in the virtual world: Users' identification with virtual communities and avatars. *Computers in Human Behavior*, *28*(5), 1663-1669.

Li, W., Cosker, D., Lv, Z., & Brown, M. (2016). Dense nonrigid ground truth for optical flow in real-world scenes. *IEEE Robotics and Automation Letters 0*(0), *12*.

Li, L., & Walsh, S. (2011). Technology uptake in Chinese EFL classes. *Language Teaching Research*, *15*(1), 99-125.

Magid, M., & Chan, L. (2012). Motivating English learners by helping them visualise their Ideal L2 Self: lessons from two motivational programmes. *Innovation in Language Learning and Teaching*, *6*(2), 113-125.

Markus, H., & Kunda, Z. (1986). Stability and malleability of the self-concept. *Journal of Personality and Social Psychology, 51*(4), 858-866.

Markus, H., & Nurius, P. (1986). Possible selves. *American Psychologist*, *41*(9), 954-969.

Markus, H., & Ruvolo, A. (1989). Possible selves: Personalized representations of goals. In L.A. Pervin (Ed.), *Goal concepts in personality and social psychology* (pp. 211-241). Hillsdale, NJ: Erlbaum.

Markus, H., & Wurf, E. (1987). The dynamic self-concept: A social psychological perspective. *Annual Review of Psychology, 38,* 299-337.

McCarthy, M., McCarten, J. & Sandiford, H. (2014). *Touchstone Level 1 Workbook 2nd Edition*. Cambridge University Press.

Nowell, L.S., Norris, J.M., White, D.E. & Moules, N.J. (2017). Thematic analysis: Striving to meet the trustworthiness criteria. *International Journal of Qualitative Methods, 16,* 1-13.

Oettingen, G., & Thorpe, J. (2006). Fantasy realization and the bridging of time. In L.J. Sanna & E.C. Chang (Eds.), *Judgments over time: The interplay of thoughts, feelings, and behaviors* (pp. 120-142). New York: Oxford University Press.

Oyserman, D., & Destin, M. (2010). Identity-based motivation: Implications for intervention. *The Counselling Psychologist*, *38*(7), 1001-1043.

Paivio, A. (1985). Cognitive and motivational functions of imagery in human performance. *Canadian Journal of Applied Sport Sciences, 10,* 228-288.

Pearson, J., & Kosslyn, S.M. (2015). The heterogeneity of mental representation: Ending the imagery debate. *PNAS, 112*(33), 0089–10092.

Reallusion. (2016). *CrazyTalk 8* [Computer software]. California, USA. Retrieved from https://www.reallusion.com/crazytalk/

Sánchez-Lozano, E., Martinez, B., Tzimiropoulos, G., & Valstar, M. (2016, October). Cascaded continuous regression for real-time incremental face tracking. In *European Conference on Computer Vision* (pp. 645-661). Springer International Publishing.

Stockwell, G. (2013). Technology and motivation in English-language teaching and learning. In E. Ushioda (Ed.) *International perspectives on motivation* (pp. 156-175). Palgrave Macmillan UK.

Szczuko P., Kostek B., & Czyżewski A. (2009). New Method for Personalization of Avatar Animation. In K.A. Cyran, S. Kozielski, J.F. Peters, U. Stańczyk & A. Wakulicz-Deja (Eds.) *Man-Machine Interactions. Advances in Intelligent and Soft Computing* (pp. 435-443). Berlin: Springer.

Taylor, S.E., Pham, L.B., Rivkin, I.D., & Armor, D.A. (1998). Harnessing the imagination: Mental simulation, self-regulation, and coping. *American Psychologist, 53*(4), 429 439.

Thies, J., Zollhöfer, M., Nießner, M., Valgaerts, L., Stamminger, M. & Theobalt, C. (2015). Real-time expression transfer for facial reenactment. *ACM Transactions on Graphics*, 34, 1-14.

Thorne, S.L. (2008). Transcultural communication in open internet enviornments and massively multiplayer online games. In S. Magnan (ed.), *Mediating discourse online* (pp. 305–27). Amsterdam: Benjamins.

Thorne, S.L., & Black, R.W. (2007). Language and literacy development in computer-mediated contexts and communities. *Annual Review of Applied Linguistics*, *27*(133–60).

Trope, Y., & Liberman, N. (2010). Construal-level theory of psychological distance. *Psychological Review, 117*(2), 440-463.

Verdet, F., & Hennebert, J. (2008, September). Impostures of talking face systems using automatic face animation. In *Biometrics: Theory, Applications and Systems, 2008. BTAS 2008. 2nd IEEE International Conference* (pp. 1-4). IEEE.

Warner, T. (2012). E-coaching systems: Convenient, anytime, anywhere, and nonhuman. *Performance Improvement*, *51*(9), 22-28.

Wei, R., & Su, J. (2012). The statistics of English in China. *English Today*, *28*(3), 10-14.

You, C.J., & Dörnyei, Z. (2016). Language learning motivation in China: Results of a large-scale stratified survey. *Applied Linguistics*, *37*(4), 495-519.

You, C.J., Dörnyei, Z., & Csizér, K. (2016). Motivation, vision, and gender: A survey of learners of English in China. *Language Learning, 66*(1), 94-123.

Figure captions

Figure 1: Process of 2D digital animation: a participant’s photograph and animated 2D output.

Figure 2: Process of 3D digital animation: a learner’s photograph, base digital output, and output manipulated to include features and accessories.

Figure 3: An outline of the facial features copied from participants' photographs.

Figure 4: Showing the before (left) and after (right) of the Facial Overlay process of one of the female protagonists in videos ‘without context’ (top) and ‘with context’ (bottom).
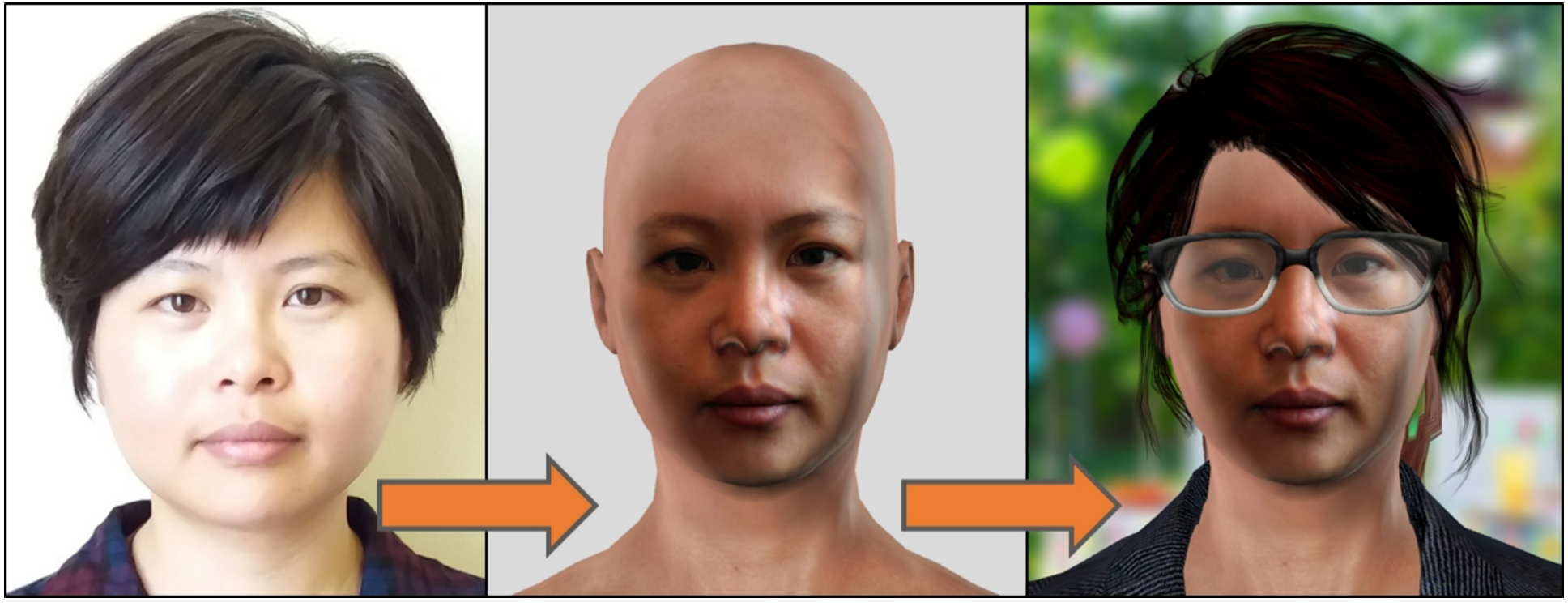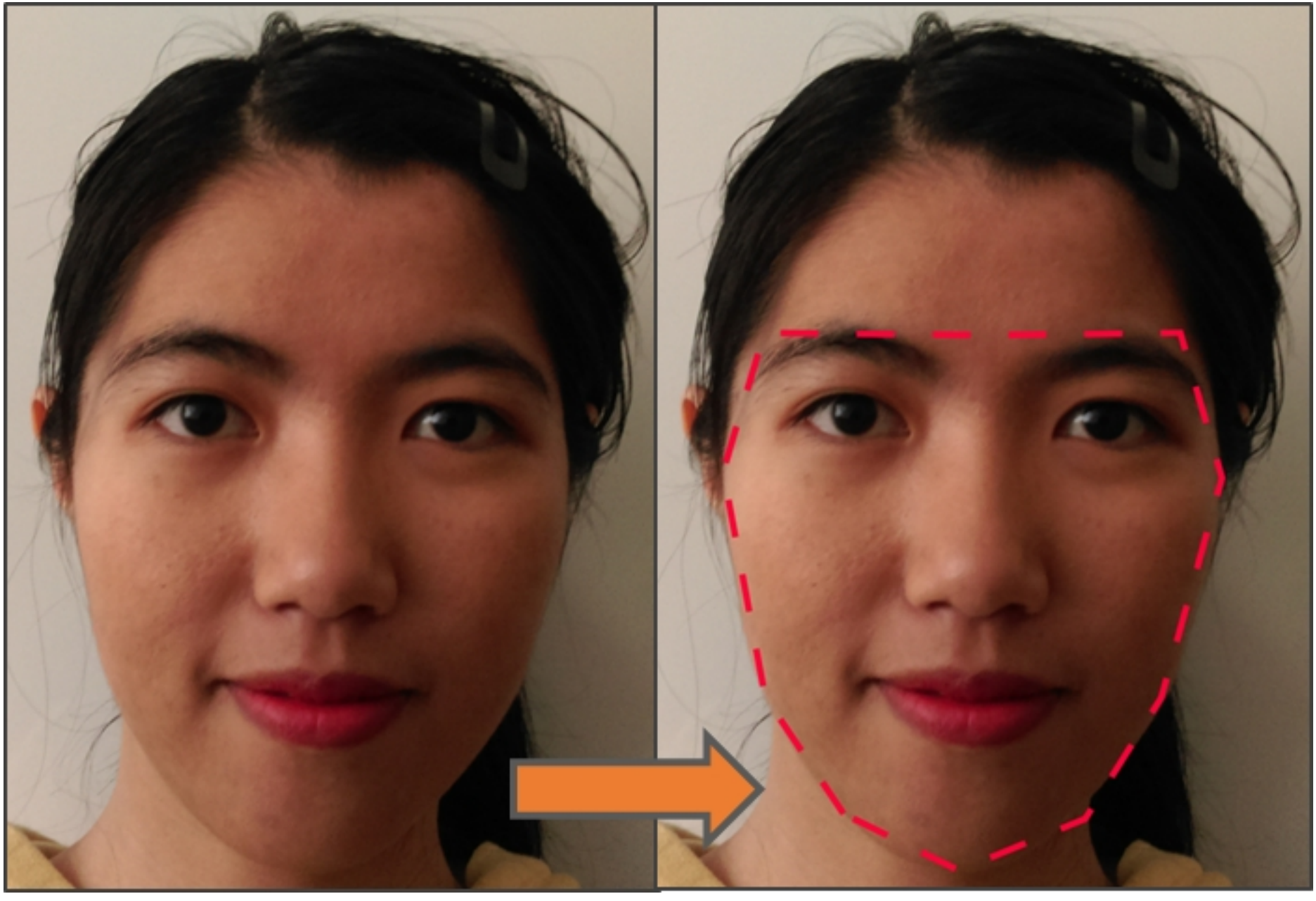
Figure 5: Facial Mask process on a female protagonist in the videos ‘without’ and ‘with context’.