

Faculté des sciences économiques,  
sociales, politiques et de communication

## Un jeu VR dédié à la prononciation de l'anglais

Une étude empirique sur l'engagement, le *flow* et  
l'immersion pour l'apprentissage par le jeu

Auteur : Alessandro Cierro  
Promoteur(s) : Thibault Philippette

Année académique 2018-2019  
Master 120 en sciences et technologies de l'information et de la  
communication



## *Remerciements*

Je tiens à remercier Alexandra Syskova, Doan Vu Duc et Gwenaël Laurent. Mon promoteur Thibault Philippette ainsi que toute l'équipe GRAAL.



# I. Introduction

## A. Contexte et démarche

Améliorer sa prononciation en langue étrangère est un défi. Si d'aucuns estimerait que la meilleure approche est de visiter sur la durée le pays de la langue ciblée, la solution n'est pas envisageable pour tout le monde. Par ailleurs, les professeurs de langue n'ont pas toujours ni le temps, ni les capacités de faire progresser leurs apprenants vers une prononciation impeccable.

Après avoir travaillé sur une solution, au sein d'une équipe pluridisciplinaire à l'UCLouvain, qui prend la forme d'une application mobile dédiée à l'apprentissage de la prononciation ; nous nous questionnons sur ce que pourrait induire sur l'amélioration de la prononciation une application plus ludique, plus immersive et plus narrative. Cette situation nous a poussé à réaliser un tel média et à le tester dans le cadre de cette étude.

## B. Définitions

Cette recherche utilise une série de concepts qu'il est nécessaire de définir et de resituer par auteur. Nous ciblerons principalement la question de l'*immersion*, de l'*engagement* et du *flow* dans un contexte d'apprentissage par le jeu.

### 1. "Serious" game et gamification

Les jeux sérieux (serious games) et la ludification (gamification) diffèrent des jeux "classiques" en ce sens que les premiers ont une visée éducative et ne servent pas exclusivement le divertissement (Davidson (2008), Hamari and Koivisto (2015)).

## 2. Concept de flow

Le sujet de cette étude est un jeu éducatif que nous avons développé. Joyner et al. (2002) décrivent ce type de jeu comme capable de combiner la concentration nécessaire à une pratique "challenging" et en même temps permettre une expérience amusante ("enjoyment experience") ; le jeu tient ainsi compte des compétences du joueur et permet un *serious play* ou un *playful work*.

Brogère (2012) décrit les critères fondamentaux qui permettent à un média d'être appelé "jeu" avant de s'intéresser au jeu sérieux.

« J'ai tendance à penser aujourd'hui que ces deux critères suffisent à analyser les situations pour en saisir la dimension ludique ou non. La présence d'une **activité de second degré** accompagné de **la décision** (en particulier collective), nous conduit de facto dans l'univers potentiel du jeu (si tous les participants sont du côté des décideurs, contrairement au spectacle où une partie des participants ne décide pas). A partir de ces deux éléments qui définissent l'espace du ludique on trouve deux caractéristiques associées (l'existence de modalités de décisions dont les règles sont un exemple et l'incertitude, l'absence de prédétermination de la fin) et une caractéristique conséquente, la frivolité ou minimisation des conséquences. Il s'agit là des trois autres critères selon mes analyses antérieures, mais que j'aurais tendance à isoler comme des conséquences des deux principaux critères. Je définirais ma proposition sous la formule 2 + 3 critères et non plus 5. Ces trois autres critères ne sont que les effets de cette configuration, cette association entre second degré et décision, caractérisant les jeux dans leur diversité :

- les jeux supposent un **système de décision** (...).
- les jeux impliquent **une incertitude**, avec les marges qui seraient la fin prévue d'avance (on sort du jeu) (...).
- les jeux impliquent une **absence de conséquence** ou minimisent les conséquences, (...). » (Nous soulignons dans la citation les critères de Brogère (2012) (p. 7-8))

Ces critères, selon Brogère, suffisent à la dénomination de jeu. Brogère poursuit avec les nuances que peut apporter le jeu sérieux.

« (...) si l'on affine l'analyse, on peut mettre en évidence le mécanisme qui permet la décision, ou la rhétorique procédurale. C'est ce que propose le jeu en tant *game*, objet technique et logiciel (*software*). Le *serious game* est pris dans une tension entre l'usage des mécanismes qui ont fait leur preuve dans les jeux vidéo (le *gameplay*) et la nécessité de disposer de procédures, de mécanismes de décision

en harmonie avec l'objectif du jeu. On voit déjà comment ce critère inscrit le *serious game* dans une tension entre logique du jeu (*gameplay*) et logique de l'apprentissage ou de l'univers de référence. C'est sans doute là que se joue la différence entre simulation et jeu de simulation, entre la volonté de reproduire procédures et mécanismes du réel et celle de valoriser des logiques ludiques issues de jeux dont le divertissement est la finalité. La logique des deux produits diffère dans la mesure où, à la base du jeu à finalité de divertissement, il y a une logique ludique (un *gameplay*) que l'on va intégrer dans une histoire, un univers fictionnel adapté à celui-ci, alors que le *serious game* va procéder à l'inverse : à l'origine il y a un univers de référence (ce qui est à mettre en valeur ou à apprendre) sur lequel on va greffer un *gameplay* plus ou moins adapté. » (Brougère (2012) (p. 9))

Dès lors, nous pouvons faire l'hypothèse d'une balance possible entre l'aspect "fun et ludique" et l'aspect "sérieux et éducatif", qui entrent en dialogue.

Csikszentmihalyi (1991) développe un concept clé, beaucoup cité dans les *Game Studies* ou *Études du jeu* et essentiel à cette recherche. Nous pouvons emprunter à Caïra (2018) sa traduction d'une définition très complète du *flow* :

« Premièrement, l'expérience se produit d'ordinaire quand nous nous confrontons à des tâches que nous avons des chances d'accomplir. Deuxièmement, nous devons pouvoir nous concentrer sur ce que nous faisons. Troisièmement et quatrièmement, la concentration est généralement possible parce que la tâche entreprise présente des objectifs clairs et nous procure un feed-back immédiat. Cinquièmement, l'on agit avec une implication profonde mais naturelle, qui libère la conscience des inquiétudes et des frustrations de la vie quotidienne. Sixièmement, des expériences agréables permettent aux gens d'éprouver un sentiment de contrôle sur leurs actions. Septièmement, le souci de soi disparaît, mais paradoxalement la conscience de soi apparaît plus forte une fois l'expérience du *flow* terminée. Enfin, le sens de la durée est altéré ; les heures passent comme des minutes et les minutes peuvent sembler durer des heures. La combinaison de tous ces éléments procure une profonde jouissance, si satisfaisante que les gens estiment qu'elle vaut la peine d'y consacrer une énergie importante.» (Csikszentmihalyi (2008) p. 49, traduction de Caïra (2018))

Le concept de *flow* a, par ailleurs, été associé à l'apprentissage, à l'accomplissement créatif ainsi qu'à la réussite académique (Csikszentmihalyi (1997), Csikszentmihalyi et al. (1993)).

### 3. Concept d'immersion

Hamari et al. (2016) décrivent le lien entre le *flow* et l'*immersion* :

« Flow theory has been a primary theoretical base for exploring the implications of learning through immersion or “being enveloped” by a virtual learning environment because the emotional composition of these experiences resemble flow and precipitate a deeper engagement with learning. Research has explicitly related the sense of “presence,” “being there,” “immersion,” or “flow” in different virtual reality interfaces with positive learning outcomes. » (e.g., Abrantes and Gouveia (2012), Fassbender et al. (2012)) » (Hamari et al. (2016), p.172)

Peu d'études questionnent la relation entre l'*immersion* et l'apprentissage dans le jeu. En excluant notre article de référence qui ne s'intéresse pas exclusivement à la question de l'*immersion*, Cheng et al. (2014) ont prouvé dans leur étude que cette dernière a pourtant une influence positive sur l'apprentissage de principes scientifiques.

Nous nous baserons sur la définition de l'*immersion* de Ermi and Mäyrä (2005) (qu'ils mettent en lien avec la notion de *gameplay*) :

« The first dimension of a gameplay experience that we distinguish is the **sensory immersion** related to the audiovisual execution of games. This is something that even those with less experience with games – like the parents of the children that were interviewed – can recognize: digital games have evolved into audiovisually impressive, three-dimensional and stereophonic worlds that surround their players in a very comprehensive manner. Large screens close to player's face and powerful sounds easily overpower the sensory information coming from the real world, and the player becomes entirely focused on the game world and its stimuli. Another form of immersion that is particularly central for games, as they are fundamentally based on interaction, is **challenge-based immersion**. This is the feeling of immersion that is at its most powerful when one is able to achieve a « satisfying balance of challenges and abilities. Challenges can be related to motor skills or mental skills such as strategic thinking or logical problem solving, but they usually involve both to some degree. In several contemporary games also the worlds, characters and story elements have become very central, even if the game would not be classifiable as an actual roleplaying game. We call this dimension of game experience in which one becomes absorbed with the stories and the world, or begins to feel for or identify with a game character, **imaginative immersion**. This is the area in which the game offers the player a chance to use her

imagination, empathise with the characters, or just enjoy the fantasy of the game. »  
(Ermi and Mäyrä (2005), pp.7-8)

Nous justifierons la réalisation de notre média au travers des trois types d'immersion (voir Figure 1) développés par Ermi and Mäyrä (2005) :

- **Sensory immersion** : représente l'aspect audiovisuel de notre média (usage de la réalité virtuelle, du tracking,...)
- **Challenge-based immersion** : décrit la balance entre les défis qu'amène le jeu et les capacités du joueur.
- **Imaginative immersion** : se définit dans les aspects narratifs et empathiques développés dans le jeu.

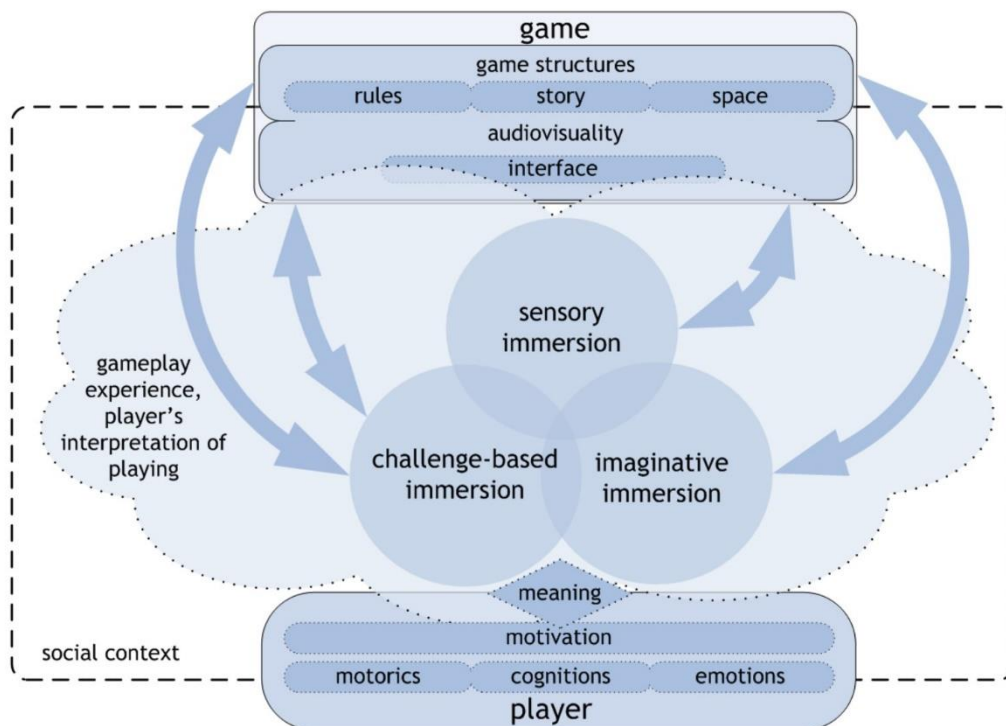


Figure 1. Modèle développé par Ermi and Mäyrä (2005) (p. 8) : les 3 types d'immersion et leur rôle dans la construction du gameplay d'une expérience.

#### 4. La perception d'apprentissage

Nous n'étudierons pas la notion d'apprentissage en elle-même car dans le cas qui nous concerne, évaluer les compétences de prononciation d'une langue étrangère est une tâche qui nécessite une expertise spécifique. Il n'existe pas non plus de test mesurant avec précision le niveau de

prononciation d'un apprenant. Enfin, la configuration de notre étude ne nous permet pas d'estimer une quelconque évolution du niveau de prononciation sur une période de temps conséquente (plusieurs semaines voire plusieurs mois de pratique) ; nous devons nous contenter d'une série d'expériences de plus ou moins 1h30.

C'est pourquoi, nous étudierons plutôt la *perception d'apprentissage* qui a été longtemps considérée (avec la perception de la compétence) comme l'un des meilleurs prédicteurs de la motivation, du comportement, et un des éléments déterminants de la réussite (*achievement expectations*) (Nicholls (1979), White (1959)).

Thomas (1980) décrit même le lien entre la perception de compétence et la perception de succès comme un élément fondamental de la motivation d'apprendre. Il indique en outre qu'à l'inverse, les étudiants qui perçoivent une forme d'incompétence se sentent inconfortables à entreprendre quoi que ce soit leur permettant hypothétiquement d'augmenter leurs compétences.

### C. Article de référence

Cette étude s'installe dans la continuité de celle de Hamari et al. (2016); nous reprendrons la même approche méthodologique, et le même modèle conceptuel. Le sujet de l'apprentissage sera cependant adapté pour correspondre à nos besoins. L'intérêt de notre démarche se fonde sur l'une des limites identifiées par Hamari et al. (2016) dans leur article : le cas de l'*immersion*. En effet, les deux médias développés dans le cadre de leur recherche n'intègrent pas les procédés immersifs pourtant potentiellement nécessaires à la confirmation de l'une de leurs hypothèses.

« In the present study immersion did not have significant relationship with perceived learning. The lack of association between immersion and perceived learning may be caused by several factors. (. . . ) Consequent studies on immersion and learning should more deeply measure immersion and the effect of different types on learning. (. . . ) Moreover, neither of the games included in this study, Spumone nor Quantum Spectre, involve a narrative driven, avatarbased, sensory or imaginative immersive experience that other games may provide. (. . . ) Further studies using 3D immersive games, such as games in virtual worlds with avatars

and a detailed environment, would be helpful to explore this possibility further. »  
(Hamari et al. (2016), p.176)

C'est donc avec ces contraintes : mondes virtuels, avatars, narration bien présente, environnement virtuel détaillé et réaliste ; que nous avons mis en place notre jeu en réalité virtuelle. Nous partageons beaucoup avec notre étude de référence : en plus de la structure générale de cette recherche (avec une attention aux mêmes contraintes méthodologiques) ; nous avons naturellement une littérature de référence similaire. Notre intérêt étant de continuer au plus près la réflexion posée par Hamari et al. (2016) en observant un média immersif dédié à l'apprentissage de la prononciation de l'anglais.

## II. Média

### A. Description du média

Le média en réalité virtuelle réalisé dans le cadre de cette étude se compose de 3 parties distinctes : une partie tutoriel, une partie jeu, et une partie feedback et score.

#### 1. Tutoriel — 1<sup>ère</sup> partie

Ce premier volet se déroulant dans le hall de la gare d'Anvers permet aux apprenants de découvrir les commandes qu'ils auront à utiliser durant toute l'expérience (voir Figure 2). Après avoir découvert et pratiqué les commandes usuelles, ils peuvent choisir leur profil de prononciation US ou UK (voir Figure 3).



Figure 2 – Tutoriel, touches et fonctionnalités possibles.

Avant de commencer l'expérience et de passer au volet suivant, nous indiquons aux apprenants qu'une flèche orange peut survenir à tout moment et ainsi notifier qu'ils passent au niveau de difficulté supérieur.

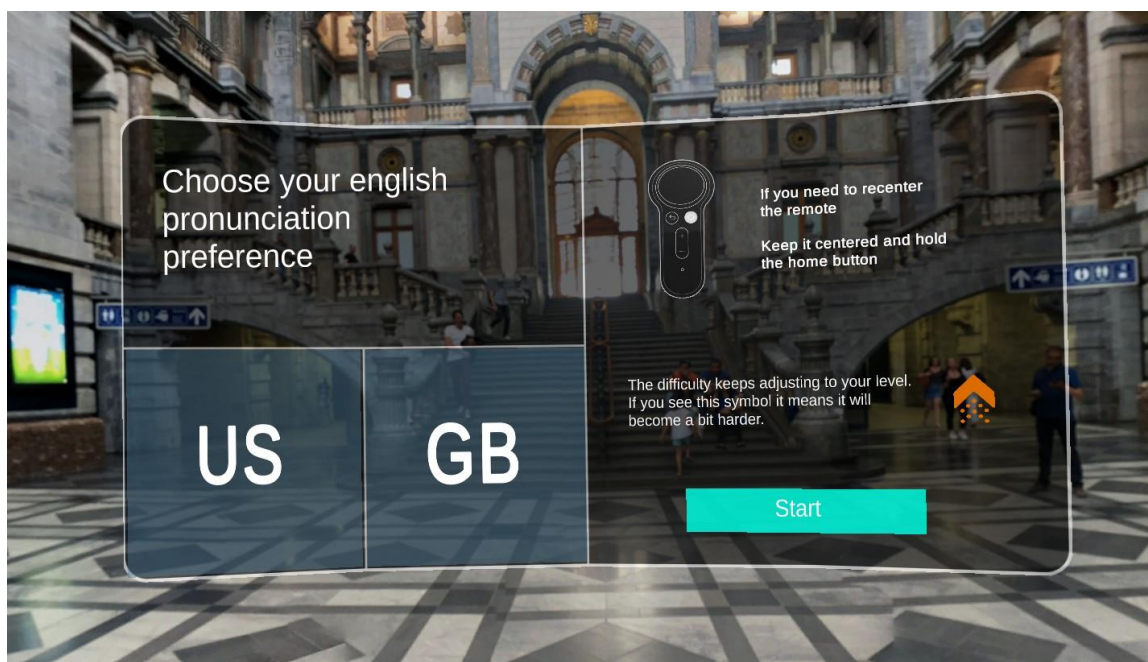


Figure 3 – Préférence de prononciation.

## 2. L'épisode du train — 2<sup>ème</sup> partie

### a. Description générale

Les apprenants seront amenés dans cette scène à parler à des avatars réalistes dans un train. Trois panneaux s'offrent à eux (voir Figure 4) : en bleu sur la gauche la phrase que vient de prononcer l'avatar, au centre inférieur ce qu'a compris la reconnaissance vocale de ce que vient de prononcer l'apprenant, et à droite les phrases que l'apprenant peut choisir de prononcer.

Si la phrase choisie est correctement prononcée, l'avatar répond et l'expérience peut se poursuivre. Dans le cas où la phrase n'est pas correcte ou suffisamment bien prononcée, l'avatar demande à l'apprenant de répéter

la phrase jusqu'à ce qu'elle soit parfaitement prononcée ou que le nivellement automatique de la difficulté permette à l'apprenant de continuer.



Figure 4 – Situation de conversation avec les personnages ; trois panneaux sont toujours présents.

En fonction du niveau de prononciation de l'apprenant, ce dernier pourra rencontrer et discuter avec 6 ou 7 personnages différents. La durée totale de l'épisode est fortement variable et dépend, par exemple, du nombre de répétitions demandées à l'apprenant, du temps qu'il met à regarder le paysage, du temps qu'il met à répondre à chaque avatar ou encore des choix narratifs opérés durant l'épisode. La durée moyenne de l'épisode est de 38 minutes.

### *b. La narration*

Un fil rouge narratif accompagne l'apprenant durant toute l'expérience. Nina, le premier personnage rencontré, nous annonce qu'elle a trouvé des tickets pour un concert de jazz sur le sol de la gare ; ces tickets ne nous appartenant pas, nous nous offrons de l'aider à trouver le propriétaire qui les a égarés. Ce qui nous amène à aller discuter avec les autres passagers du wagon (voir Figure 5 et Figure 6).

Nous pouvons ainsi discuter avec Nina : une étudiante qui se rend à un examen d'admission ; Marcus : un homme d'une quarantaine d'années, aveugle et passionné de jazz ; Samantha et Danny : deux enfants bruyants avec lesquels nous allons jouer au jeu des devinettes ; Helen : une femme âgée et taquine qui voyage pour la première fois ; Min : le contrôleur du train ; et - si le niveau de l'apprenant est suffisamment élevé - Chris, grâce à qui nous apprenons à qui appartiennent ces tickets.

La narration interactive assure à l'apprenant de jouer à plusieurs reprises le média et de découvrir de nouvelles phrases, de nouvelles fins et d'avoir des discussions différentes avec chaque personnage.

La progression dans la narration en arborescence s'opère de deux manières : en fonction des phrases choisies par l'apprenant, et en fonction de son niveau de prononciation. Le média reconnaît 8 profils de difficultés de prononciation et dirige l'apprenant au travers de l'histoire correspondante tout en prenant en considération ses choix de phrases en réponse.

Très concrètement, le média offre une multitude de variantes narratives, 26 fins différentes et un personnage supplémentaire si le niveau de l'apprenant le permet. Son niveau ainsi que ses choix influencent l'expérience dans son ensemble.

Voici certaines des fins possibles : nous retrouvons le propriétaire des tickets, nous offrons les tickets à l'un des passagers, nous gardons les tickets pour nous, il nous est proposé d'accompagner l'un des passagers au concert de jazz et de faire plus ample connaissance.



Figure 5 – De gauche à droite : Nina, Markus et les enfants : Danny et Samantha.



Figure 6 – De gauche à droite : Helen, Min et Chris.

### *c. Challenge et niveau de difficulté*

Le média adapte la difficulté de l'expérience selon le niveau de l'apprenant. Ce niveau est adapté dynamiquement en permanence ; le but étant de permettre à l'apprenant de se sentir challengé tout au long de l'épisode et de progresser depuis son propre niveau.

Ce nivelage lui permet d'avoir une expérience au plus proche de son niveau. Les éléments qui varient en fonction des 8 profils de niveau sont les suivants :

- Le nombre de personnages ;
- La marge d'erreur tolérée (plus le niveau de l'apprenant est diagnostiqué comme fort, moins d'erreurs lui sont permises pour avancer dans la narration) ;
- La complexité des phrases proposées à l'apprenant ;
- La précision de l'analyse de la reconnaissance vocale (si le niveau est plus élevé, plus d'éléments de prononciation sont considérés : la prosodie, les contractions, les lettres aspirées ou silencieuses...) ;
- Le nombre de phrases proposées à l'apprenant ;
- Le nombre d'interactions avec chaque personnage ;
- La durée totale de l'épisode.

Deux éléments font état de l'adaptation de la difficulté : le symbole de la flèche orange dans le cas où le niveau augmente ; au contraire, si une phrase n'est pas correctement prononcée à plusieurs reprises, le niveau peut diminuer, permettant à l'apprenant de progresser malgré tout.

### 3. Feedback et score — 3<sup>ème</sup> partie

La dernière partie de l'expérience offre à l'apprenant un feedback de toutes les phrases prononcées ainsi que son score général et son profil de niveau (nombre de "Level Up" remis sur 20).

#### *a. Score total et Level Up*

Le score et le Level Up doivent se lire ensemble : si l'apprenant a obtenu 88% et 12 Level Up, il faut comprendre que le niveau 12 est presque complètement acquis et qu'il lui reste certaines améliorations nécessaires pour atteindre le niveau 13. Le niveau 12 se veut être proche des *Niveaux de langues européens* (CECRL) B1-B2.

### *b. Feedback*

Chacune des phrases sont reprises et classées par conversation. Une appréciation (sur 3 étoiles) est proposée pour chaque phrase ; et s'affiche en rouge ce qui aurait dû être mieux prononcé pour atteindre le niveau suivant.

## **B. Conception du média**

### **1. Contexte**

Ce média en réalité virtuelle a été réalisé dans le prolongement du projet Spin-off GRAAL du laboratoire Pixel & Interaction lab à l'UCLouvain. Le projet GRAAL (pour Guidage en Réalité Augmentée de l'Apprentissage des Langues) a pour but de proposer une solution d'apprentissage et d'amélioration de la prononciation en langue étrangère. Cette spécificité de l'apprentissage de langue est souvent écartée du parcours pédagogique classique, pour des raisons multiples : absence de professeurs natifs, pas toujours de temps dédié spécifiquement à la prononciation, pas de légitimité (tous les professeurs n'ont pas appris à enseigner la prononciation),...

Le projet Spin-off met en place une application d'entraînement intensif de la prononciation qui repose sur 2 piliers : une analyse vocale (speech processing) réalisée dans les laboratoires et se ciblant spécifiquement sur les besoins des apprenants ; et un parcours d'apprentissage finement pondéré et construit par un ensemble de professeurs de langues, de linguistes et de pédagogues.

Cependant, si GRAAL est un projet complet, il ne bénéficie pas de dimension immersive particulière. Ainsi, pour isoler la variable d'immersion au sein de l'apprentissage de la prononciation, nous avons réalisé un média narratif en réalité virtuelle. Le but étant de cibler les avantages pédagogiques que peut revêtir l'immersion VR.

Nous n'avons pas eu accès aux ressources pédagogiques ou techniques développées au sein du projet GRAAL pour la réalisation de notre média. La solution réalisée est autonome et embrasse les mêmes objectifs que GRAAL

pour assurer au projet Spin-off d'avoir une vision claire des effets de l'immersion dans un cadre d'apprentissage similaire au sien.

## 2. Réalisation

### *a. Matériel*

Nous avons fait le choix de réaliser une application fonctionnelle sur smartphone ; l'aspect nomade de notre média aura grandement facilité les tests. Nous avons utilisé le casque de réalité virtuelle signé Oculus et Samsung : le Gear VR. Il possède également une télécommande et assure une expérience immersive proche de celle des grands casques de réalité virtuelle basés sur ordinateur. Le smartphone utilisé avec le Gear VR fut un Samsung S10.

Les deux principales différences entre un casque VR nomade basé sur un smartphone et un casque VR basé sur ordinateur sont :

- La qualité des rendus visuels (et la puissance de calcul).
- L'absence de capteurs de suivi de mouvements permettant à l'utilisateur de se déplacer librement dans un espace déterminé et balisé.

Durant la réalisation du média, ces deux éléments essentiels à l'immersion ont donc été considérés. Nous avons décidé d'utiliser un smartphone puissant (CPU : Exynos 9820 et 6 gigas de RAM) avec une résolution élevée (1440 x 3040 pixels avec une densité de ~550 ppi) assurant une résolution par œil de 1280 x 1440 pixels (comparable à l'Oculus Go). Chaque scène a été optimisée au mieux pour assurer un rendu fluide de l'expérience en réalité virtuelle.

Pour ce qui est du suivi de mouvement sur smartphone, nous proposons une solution novatrice que nous développerons dans le point qui suit.

### *b. Unity3D*

Le moteur de jeu Unity3D a été utilisé pour la réalisation de notre média. Plusieurs dimensions techniques peuvent être développées en détail.

### *c. Modélisation 3D*

Il a été nécessaire de modéliser certains éléments pour permettre une expérience immersive réaliste complète (*sensory immersion*). Le programme de modélisation Autodesk Maya a permis la création du train dans son ensemble ; les avatars ont été réalisés sur MakeHuman ; les animations des personnages construites sur Maya et optimisées sur Unity ; et enfin deux images 360° ont été utilisées : un paysage d'Arizona (pour le paysage visible depuis le train) et une deuxième que nous avons réalisée à la gare d'Anvers (pour le tutoriel et le feedback final).

L'ensemble de la modélisation a dû tenir compte des capacités limitées de rendu des smartphones. Dès lors, nous avons optimisé chaque modèle 3D afin qu'il possède un nombre minimum de polygones : ce balancement entre "Low- Poly" et modèle ultra réaliste avait comme objectif d'assurer l'immersion le plus efficacement possible (nous avons modélisé des avatars avec le niveau de détail correspondant au deuxième modèle sur les 3 représentés en Figure 7).

### *d. Technologie vocale (TTS)*

Il était capital de doter chaque avatar d'une voix la plus humaine et cohérente possible ; ceci dit, il n'est pas aisé de trouver des voix générées par ordinateur suffisamment convaincantes. Nous avons utilisé les nouvelles voix de synthèse Wavenet de Google pour Nina, Markus, Min et Chris, tandis que la voix d'Helen et celles des deux enfants furent réalisées à partir des modules d'Acapela group.

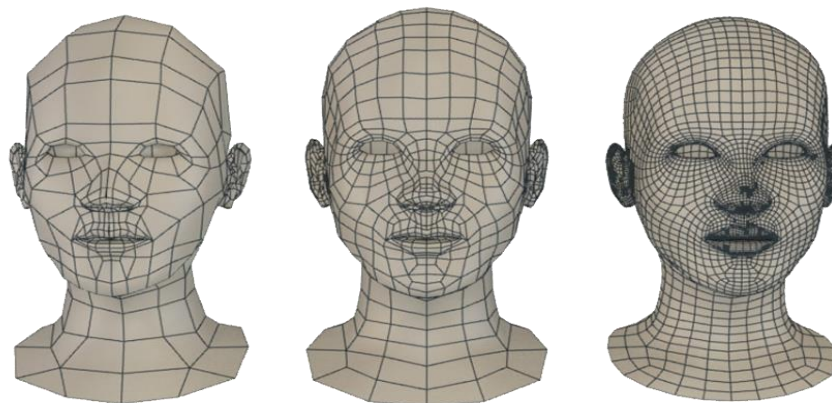


Figure 7 – Un visage de femme modélisé en 3D avec une variation polygone.  
De gauche à droite Low-poly avec 570 polygones, Mid-Poly avec 1102 polygones et High-Poly 9180 polygones - Modèle 3D de LBG4 disponible sur [turbosquid.com/3d-models/female-head-bust-pack-3d-model/683303](https://www.turbosquid.com/3d-models/female-head-bust-pack-3d-model/683303)

#### e. Technologie de reconnaissance vocale (STT)

Le fondement pédagogique de notre média repose sur sa capacité à différencier une phrase correcte et une phrase erronée ; à reconnaître le niveau de prononciation de l'apprenant de manière dynamique ; et à modifier le niveau de difficulté en conséquence. Ou encore, à comprendre parmi les phrases proposées à l'apprenant laquelle a été sélectionnée par ce dernier, même si elle n'a pas été parfaitement prononcée. L'adaptation du niveau de l'expérience a pour visée de permettre dans une certaine mesure le *challenge-based immersion*.

N'étant pas expert en analyse du signal et de la parole, nous avons fait le choix d'utiliser les services de reconnaissance vocale de Google. Le protocole de Google nous permet de récupérer trois informations avec lesquelles nous avons construit notre algorithme :

- La transcription de la parole (Speech To Text ou STT).
- Un degré de certitude indiquant la fiabilité du STT.
- Des variantes de la transcription avec des degrés de certitude inférieurs.

Prenons un exemple concret servant d'illustration. L'apprenant a le choix de prononcer l'une ou l'autre de ces deux réponses :

1. Yes, it's my very first time. (phonétiquement : [jɛs, Its maɪ 'vɛri fɜst taɪm]).
2. No, actually I'm used to it. (phonétiquement : [n , 'æk li aɪm juzd tu It]).

Admettons que l'apprenant choisisse la première phrase et la prononce avec quelques erreurs en disant : Yes it is my first t/I/m (phonétiquement : [jɛs, It Iz maɪ fɜst tɪm]). Les 4 différences avec la phrase originale sont : l'absence de la virgule, la contraction *it's* dans sa forme *it is*, *very* a été oublié et le mot *time* qui n'a pas été bien prononcé car le son /aɪ/ a été remplacée par le son /I/.

Voici les transcriptions que renvoie la technologie de reconnaissance vocale de Google du plus au moins probable, basés sur leur degré de certitude :

1. Yes it is my first Tim – 98% de certitude.
2. Yes it is my first tim – 81% de certitude.
3. Yes it is my first team – 73% de certitude.
4. Yes it is my first teem – 32% de certitude.
5. Yes it's my first Tim – 25% de certitude.
6. Yes it's my first tim – 16% de certitude.
7. Yes it's my first team – 1% de certitude.

Grâce à une comparaison statistique des ressemblances, il est aisé de déterminer quelle phrase l'apprenant a choisi de prononcer, même si cette dernière n'est pas parfaitement correcte. A partir de ces éléments, il nous est possible de déduire les erreurs de prononciation des apprenants en comparant les transcriptions phonétiques obtenues aux phrases attendues.

Il nous faut encore déterminer quel est le seuil d'erreur acceptable pour progresser dans le média. Dans le cas illustré ici, *it's* devenu *it is* n'est pas considéré comme une erreur en anglais et donc ne peut être retenu ; il en est de même pour le cas de la virgule ou de tout autre élément de ponctuation qui

n'est jamais transcrit par la reconnaissance vocale de Google. Pour ce qui est du *very* qui a été omis, il est possible de discuter son importance relative en ajoutant une analyse sémantique à la phrase prononcée. Si la présence, le remplacement ou l'absence d'un mot ou d'un groupe de mots ne modifient pas le sens sémantique de la phrase, nous avons fait le choix d'accepter ces variations dans une certaine mesure. En conséquence, nous n'avons limité l'acceptation des variations sémantiques qu'aux apprenants avec un niveau de prononciation détecté supérieur à 14/20 (B2+), correspondant à un niveau moyen-élevé ; pour assurer aux débutants la rigueur et la précision nécessaires à leur apprentissage.

En revanche, la dernière erreur est d'importance pour un dispositif ciblant l'amélioration de la prononciation. Si un mot n'est pas correctement prononcé, la progression dans le média n'est pas possible et la phrase est systématiquement à répéter par l'utilisateur.

En fonction du niveau de l'apprenant, certaines erreurs sont considérées ou non. Une erreur comme [taIm] vs [tIm] n'est jamais acceptée car elle intervient sur la compréhension même de la phrase. D'autres erreurs n'intervenant pas sur le sens, comme une intonation non suffisamment appuyée, le manque de "sentence stress", des phonèmes qui ne sont pas suffisamment voisés,... ne sont des éléments considérés qu'à partir des niveaux supérieurs (à partir du niveau 13/20).

Le niveau est calculé et adapté en permanence et modifie directement l'expérience. Afin de déterminer le niveau de difficulté, nous avons au préalable cerné quelles erreurs appartenaient à quel niveau de difficulté (ou profil de prononciation). En repérant des erreurs ou des succès de différents types, nous pouvons placer de continuellement le niveau de l'apprenant. Notre algorithme prend en considération la répétition éventuelle de mêmes erreurs ou leur disparition progressive, combien de répétitions sont nécessaires à l'apprenant pour ne plus faire l'erreur, et propose des phrases possédant le même type d'erreur pour garder une difficulté grandissante tout au long de l'expérience.

Par exemple, prenons le cas empirique d'un apprenant ayant eu des difficultés à prononcer le son -r en anglais durant l'expérience. Toutes les phrases qui lui ont été proposées par la suite possédaient un nombre important de-r afin de lui permettre de pratiquer et d'essayer de dépasser sa difficulté. Si une difficulté est persistante, l'algorithme modifie le niveau de difficulté du média.

Lorsque le niveau est très faible (inférieur à 5/20 correspondant à A0-A1), l'expérience se modifie légèrement, et permet à l'apprenant d'entendre les phrases à prononcer avant de s'y essayer. L'avis des pédagogo-linguistes travaillant sur le projet GRAAL était tranché : si le niveau de l'apprenant est faible il est utile de lui faire entendre à l'avance la phrase à prononcer ; à l'inverse des apprenants aux niveaux moyens qui n'apprendront rien durablement en ne faisant que répéter une phrase (sans chercher au préalable et par eux-mêmes la prononciation correcte).

#### *f. Suivi du mouvement*

Nous essayons dans cette étude d'observer particulièrement l'apport de l'immersion dans un média éducatif. Y parvenir nous demandait de proposer une solution la plus immersive possible.

L'utilisation du smartphone risquait d'être responsable d'une réduction de l'immersion si les apprenants n'étaient limités qu'à un seul point de vue ; non pas que le déplacement soit spécifiquement nécessaire dans notre média, mais il n'en reste pas moins un facteur immersif considérable. A plusieurs reprises, certains de nos testeurs se sont spontanément levés, déplacés dans le wagon ou encore ont cherché à repérer les autres personnages.

La technologie aujourd'hui ne permet pas à la solution VR sur smartphone de bénéficier d'un suivi du mouvement ; les casques VR sur ordinateur utilisent une série de capteurs à placer dans la salle accueillant l'utilisation de la réalité virtuelle. Ces capteurs fonctionnent en calculant précisément la position de l'utilisateur à l'aide de la puissance de calcul de l'ordinateur sur lequel ils sont branchés.

Nous avons développé une solution alternative pour tout de même assurer le suivi des mouvements ; en exploitant l'appareil photo des smartphones qui, grâce à une série de protocoles en réalité augmentée, sont capables de repérer des plans et des arrêtes (murs, sols, objets...) entourant l'utilisateur, et d'en mesurer la distance : si l'utilisateur se déplace dans son salon, il se rapprochera ou s'éloignera des plans et arrêtes repérés. Nous reprenons ces mesures pour en calculer la distance parcourue et, directement, la retranscrire en mouvement dans la réalité virtuelle. Ces protocoles sont peu coûteux en calcul, et permettent une précision subcentimétrique. L'intérêt d'une technologie comme celle-ci est de pouvoir se passer de capteurs et d'abondants calculs tout en se libérant de la contrainte d'espace dédié. Il est possible d'utiliser cette solution à l'intérieur tout comme à l'extérieur sans aucune différence.

*g. Narration en arborescence*

Si la *sensory immersion* et la *challenge-based immersion* ont été traitées précédemment, il était important de considérer l'*imaginative immersion* Ermi and Mäyrä (2005). Nous avons donc écrit des dialogues fournissant une personnalité forte aux personnages. Il était important que l'utilisateur puisse croire aux personnages, à leur handicap, à leurs volontés, à leurs ressentis,... pour qu'il soit possible d'initier une forme d'empathie à leur égard.

Évidemment, se plonger dans une histoire nécessite un certain temps, et en 25 minutes d'expérience, il est très difficile d'assurer ce genre d'immersion. C'est pourquoi, pour amener l'utilisateur à faire des choix qui lui ressemblent, nous avons écrit une narration non-linéaire interactive à l'image des "livres dont vous êtes le héros" dans lesquels l'histoire se modifie en fonction du choix pris par le lecteur.

Cette narration interactive nécessite d'écrire tous les dialogues en arborescence, de prévoir un déroulement différent pour chaque personnage et de construire une multitude de fins. Elle permet à l'utilisateur de réaliser à plusieurs reprises l'épisode du train et d'apprendre de nouveaux éléments narratifs pour chaque personnage en plus des 26 différents dénouements élaborés.

## C. Limites et difficultés

La réalisation du média a soulevé un ensemble de défis et plusieurs difficultés plus ou moins contraignantes. Essentiellement, les grandes limites de notre média proviennent du fondement technologique sur lequel repose le principe de la reconnaissance vocale. La reconnaissance vocale est une technologie qui se base notamment sur un modèle linguistique propre à chaque langue et une série d'évaluations statistiques en temps réel. Cette approche ne permet pas de discriminer l'entière des mots/sons prononcés par l'apprenant car le modèle linguistique de la langue peut prendre le dessus et indiquer la solution correcte la plus probable et non la solution erronée prononcée.

Prenons, pour illustrer, le cas de l'apprenant qui veut dire : Yes, it's my very first time. Et prononce : Yes, it's my vary fist time. La reconnaissance vocale assume qu'il y a des erreurs dans la reconnaissance. Elle cherche alors le plus probable statistiquement et transforme une première fois la phrase prononcée en : Yes, it's my vary first time ; et une deuxième fois en : Yes, it's my very first time. Ces modifications s'opèrent car fist et vary n'ont pas de sens d'après le modèle de langue utilisé.

Or dans le cadre d'une analyse précise de la prononciation, il est nécessaire de pouvoir se rendre compte des erreurs des apprenants et leur proposer des feedbacks et conseils pertinents. Cette limite de la reconnaissance vocale a été atténuée dans une certaine mesure pour cette recherche par une analyse de l'ensemble des variations de transcription fournies par le service de Speech To Text. Nous avons observé si des variations existaient et se répétaient pour isoler le ou les sons problématiques. Dans l'exemple décrit précédemment, le mot *time* n'a pas été correctement prononcé, et des variations plus ou moins probables ont alors été proposées par la technologie (Tim - 98%, team 73%, teem 32%) ; il est possible de déduire le son /I/ qui fait défaut.

Cette déduction n'est malheureusement pas à 100% fiable car elle repose toujours sur des mots reconnus par le modèle de langue (ce qui est le

cas des variations captées : *Tim* est un prénom, *team* veut dire équipe et *teem* signifie pulluler, abonder). Par conséquent, si l'apprenant prononce un son ou un mot qui n'est pas existant dans le modèle de la langue cible, la reconnaissance vocale indiquera le mot le plus probable (en termes de sens) et le plus proche (phonologiquement) possible.

La précision qu'offre cette technologie ne suffit pas à une pratique rigoureuse et approfondie de la prononciation. C'est pourquoi, dans l'équipe GRAAL, deux ingénieurs en traitement du signal construisent une technologie propre au projet avec une analyse rigoureuse de chaque son prononcé. Une analyse complète comme l'offre GRAAL permet un feedback riche pédagogiquement avec un retour *son par son*, et non pas *mot par mot* comme dans notre média. A chaque fois qu'il fut possible de discriminer le son problématique, nous l'avons mis en exergue et en avons tenu compte dans l'évaluation permanente et dynamique du niveau de l'apprenant. Plus la technologie utilisée ici est précise, plus le média sera efficace et pédagogiquement pertinent.

Une deuxième limite à laquelle nous avons dû faire face est l'arrêt soudain des services STT du *cloud* de Google. Nous avons effectivement commencé le développement du média avec les services *cloud* de Google, qui nous permettaient d'enregistrer chaque phrase prononcée par l'apprenant, d'en obtenir un feedback assez précis, et de bénéficier d'une analyse en-dehors de la plateforme utilisée (les résultats restent les mêmes sur ordinateur, Android et iOS). En contrepartie, une connexion internet rapide était nécessaire, et le temps d'analyse était de l'ordre de 1,8 seconde (ce qui est conséquent). En outre, nous étions limité à une analyse générale et non dirigée vers un anglais spécifique (UK ou US).

Cependant, les conditions d'utilisation des services *cloud* de Google ont été mises à jour entre-temps, et plus aucune de nos versions du média n'ont été fonctionnelles. Dès lors, les séquences audio n'étant plus analysées par le *cloud* de Google, il nous a fallu chercher une autre solution fonctionnelle et tout aussi efficace. Nous avons donc élaboré un protocole différent permettant d'obtenir une analyse similaire : des fonctionnalités hors-ligne internes à Android. Cela nous a fait gagner en précision et en vitesse d'analyse – ces

fonctionnalités se sont donc avérées plus performantes que le système précédent (nous sommes passé de 1,8 seconde à 0,2 seconde), et nous pouvons désormais, depuis Android, stipuler quel accent nous souhaitons pratiquer (ce qui augmente la précision de l'analyse). En revanche, nous n'avons plus la possibilité de sauvegarder l'enregistrement de toutes les phrases prononcées, et nous ne bénéficions plus des statistiques générales générées par Google Cloud.

Dans une version optimale de notre média, nous utiliserions une technologie reposant sur une analyse sans modèle de langue contraignant : cette technologie serait en mesure de capter tous les types d'erreurs de prononciation, ce qui nous permettra d'en faire un tableau d'objectifs et de confirmer avec certitude quel objectif est atteint par l'apprenant ou ce qui reste à améliorer. Nous serions en mesure d'étoffer le feedback obtenu pour chaque phrase au regard de ces objectifs de prononciation quantifiables. Enfin, nous proposerions à l'apprenant de se réécouter si nécessaire ; et directement, depuis le feedback, proposer une nouvelle tentative avec une analyse complète en temps réel.

D'autres d'améliorations notables seront à considérer dans une version optimale :

- un enregistrement de voix humaine au lieu des voix de synthèse actuellement utilisées ;
- une amélioration des animations des personnages, qui pourraient mobiliser leurs mains pendant qu'ils parlent ;
- une amélioration des expressions faciales pour le moment statiques mais qui pourraient être animées en fonction des émotions à exprimer ;
- une animation des lèvres des personnages lorsqu'ils parlent ;
- davantage de variations narratives proposées ;
- une indication en temps réel des sons qui ne sont pas bien prononcés ;
- un mode multi-joueurs ;
- une conception de plus de 8 profils de difficulté ;

- une version basée sur une intelligence artificielle ne nécessitant plus de phrases préprogrammées, ce qui permettrait une discussion libre avec un avatar et une analyse de la prononciation à la volée.

### III. Méthode

#### A. Participants et procédures

Notre recherche s'est déroulée en plusieurs étapes : accueil du participant : remerciements et explication sur le déroulement de l'expérience qui suit.

- premier questionnaire démographique (quanti) : voir Annexe A ;
- mise en place du dispositif (casque) de réalité virtuelle et autonomie de l'utilisateur, accompagnement en cas de problème technique ;
- deuxième questionnaire (quanti) (Q2), voir Annexe B;
- phase qualitative libre : feedback, retours et commentaires, améliorations possibles, voir Annexe C ;
- fin de l'expérience.

La durée moyenne totale de l'expérience est de 85 minutes par participant. Les différents prototypes qui ont permis la réalisation de notre média ont été testés et corrigés sur base des retours de 8 bêta-testeurs.

Notre étude compte 54 participants aux profils démographiques variés : 29 femmes et 25 hommes ; 48% des participants ont entre 18 et 24 ans, 21% entre 25 et 34 ans, et 23% ont plus 44 ans ; 73% des personnes interrogées sont nées en Belgique, 5% en France, et sinon 4 sont africains, et un participant est ressortissant d'Asie centrale. Plus de 86% des participants ont le français comme langue maternelle ; près de 31% n'emploient que rarement l'anglais oralement (moins d'une fois par mois), et 15% le parlent tous les jours. Ils estiment en moyenne leur niveau de prononciation plus faible que leur niveau d'anglais général (connaissance en vocabulaire, grammaire, conjugaison). 61,5% des participants n'utilisent pas d'outils technologiques servant l'apprentissage des langues ; seuls 3 participants en font usage quotidiennement. Enfin, la réalité virtuelle fut pour 73% des personnes interrogées une totale nouveauté, tandis que 27% en ont déjà très occasionnellement fait l'expérience.

Les critères de notre échantillon étaient larges : la seule contrainte était de ne pas être parfaitement bilingue (anglais comme langue maternelle, ou spécialiste de la langue).

En conséquence, nous avons exclu de nos données 2 participants, deux hommes ayant pour première langue l'anglais et n'étant dès lors pas considérés comme pertinents dans l'analyse de nos résultats. Cependant, leurs retours positifs furent intéressants pour mesurer les limites de notre média, nous assurer de l'efficacité de notre évaluation du niveau de prononciation, et challenger la technologie de reconnaissance vocale.

Les 52 participants restants ont testé notre média dans un espace calme dans lequel nous étions seuls tout le long de l'expérience. Nos expérimentations ont pris place autant à Louvain-la-Neuve qu'à Bruxelles. Aucune interruption, bug technique ou perte de données ne sont à déplorer.

Une fois l'expérience en réalité virtuelle aboutie, les participants ont été amenés à répondre à un questionnaire psychométrique composé de 16 questions. Ce questionnaire consiste en la récolte subjective et par auto-déclaration des niveaux d'*engagement*, de *challenge*, de *compétence*, d'*immersion* ressentis et estimés par le participant ainsi qu'en son degré de perception d'apprentissage ou d'amélioration de sa prononciation durant l'expérience. Nous avons développé plusieurs questions pour chaque variable observée, posées aléatoirement et récoltées sur une échelle de Likert (de 1 à 5).

## B. Validité

Nous avons choisi d'exploiter la technique d'analyse statistique SEM pour *Structural Equation Modeling* (Anderson and Gerbing (1988), Hair et al. (2009), Hair et al. (1998)). Cette approche offre la possibilité de concevoir des modèles multivariés avec une analyse sur plusieurs niveaux : cela permet de mesurer des effets induits par de multiples facteurs. En plus d'être particulièrement adaptée aux questionnaires psychométriques, la méthode SEM permet une analyse de modèles plus complexes que ce que peut gérer une régression linéaire traditionnelle (Bagozzi and Yi (2012)).

Nous avons effectué nos tests statistiques via le programme SmartPLS 3.0. Les deux tests spécifiques opérés ont été : d'une part *PLS-SEM* pour Partial Least Squares Structural Equation Modeling, considéré adéquat pour les recherches psychométriques comme la nôtre (Ringle et al. (2005)) ; d'autre part *Bootstrapping* pour un test de significativité, ce qui permettait d'obtenir la valeur-t pour chacune de nos hypothèses.

Selon certaines études, notre échantillon est un petit peu faible pour nous assurer d'avoir des données viables et significatives. D'après Anderson and Gerbing (1988), l'échantillon minimum pour toute étude faisant usage de cette approche statistique doit être composé d'au moins 150 individus ; pour une deuxième étude, l'échantillon idéal serait composé de 5 individus par variable observée, ce qui demande donc un minimum de 70 individus dans notre cas. Enfin, pour Chin et al. (2003), le minimum satisfaisant pour un modèle SEM se composerait d'un échantillon équivalent au plus grand nombre de variables qui pointent sur la même variable à l'intérieur du modèle, multiplié par 10 (ici : Challenge, Compétence, Immersion et Engagement pointent tous vers Perception d'apprentissage). Dans notre étude, 40 individus serait l'extrême minimum.

Notre recherche se place dans la continuité de celle de Hamari et al. (2016) : nous avons repris le même schéma d'étude (voir Figure 8) en nous centrant spécifiquement sur la question de l'*immersion*. Malgré tout, la structure générale de son modèle a été respectée pour assurer, par la suite, une comparaison de nos résultats.

Ajoutons que nous portons une importance toute particulière à H1 qui cible les effets de l'immersion sur la perception d'apprentissage. L'idée étant de comparer nos résultats à ceux de la littérature avec un média en réalité virtuelle spécifiquement dédié à l'*immersion*.

## IV. Résultats

### A. Corrections et adaptations du modèle

Nos résultats ont très rapidement montré quelques incohérences ou un manque de consistance dans nos données qu'il nous a fallu corriger.

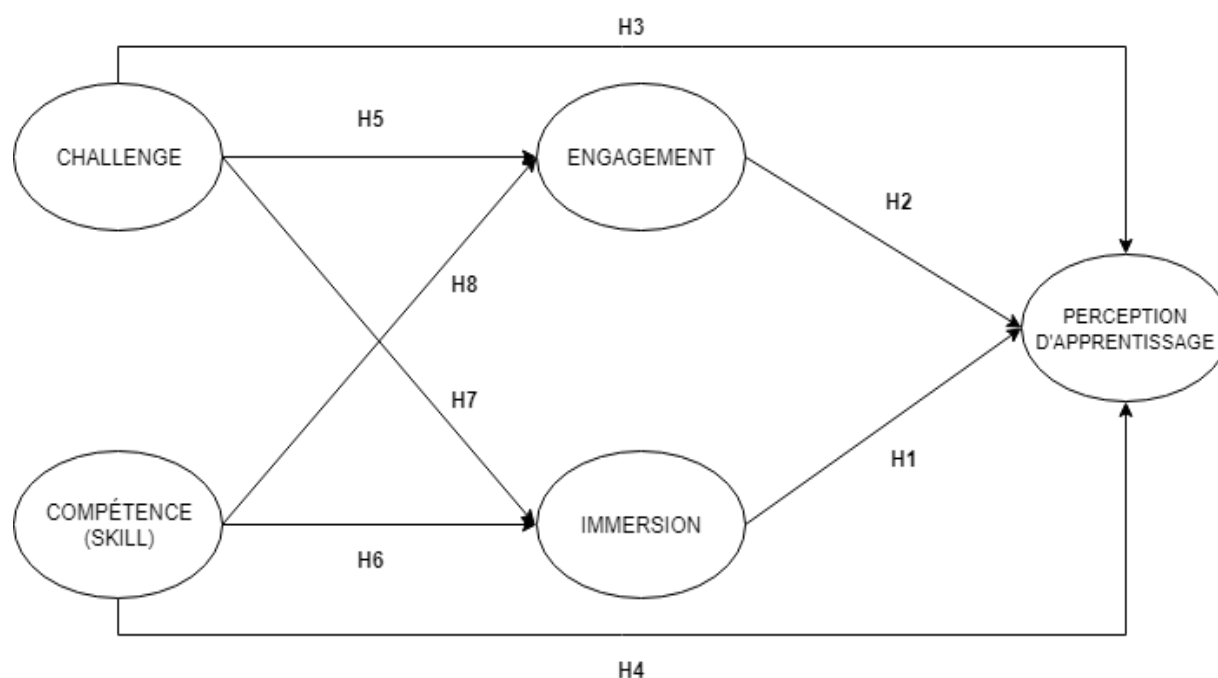


Figure 8 – Hypothèses de recherche :

H1 : Immersion aura un effet direct positif sur la perception d'apprentissage.

H2 : Engagement aura un effet direct positif sur la perception d'apprentissage. H3 : Challenge aura un effet direct positif sur la perception d'apprentissage.

H4 : Compétence aura un effet direct positif sur la perception d'apprentissage.

H5 : Les effets de Challenge sur la perception d'apprentissage seront partiellement médiés par Engagement.

H6 : Les effets de Compétence sur la perception d'apprentissage seront partiellement médiés par Immersion.

H7 : Challenge aura un effet direct positif sur Immersion. H8 : Compétence aura un effet direct positif sur Engagement.

Concrètement, nous avons observé que les questions "Immersion 3" et "Skill 2" n'étaient pas du tout corrélées aux variables respectives Immersion et Skill. Ces deux questions (respectivement : *A quel point désiriez-vous découvrir à qui appartenait les places de concert égarées ?* et *Avez-vous choisi le plus souvent les phrases qui vous semblaient les plus difficiles parmi les propositions ?*) ne ciblent pas tout à fait ce que nous essayons d'observer.

En effet, la première question désigne l'intérêt du fil rouge narratif plutôt que celui de l'*immersion* ; pour ce qui est de la seconde question, nous pensons que les testeurs allaient potentiellement se mettre en difficulté lorsqu'ils se sentaient à l'aise avec le média, tout comme nous espérions que les testeurs avec un niveau plus faible risquaient de choisir les phrases les plus courtes et les plus simples pour éviter d'être contraints de répéter jusqu'à la bonne prononciation atteinte. En pratique, les testeurs ont choisi les phrases correspondant le plus à ce qu'ils avaient envie de dire ou d'expérimenter d'un point de vue narratif plutôt qu'en fonction d'une considération de la difficulté intrinsèque à la phrase. Ces deux questions n'étant plus pertinentes, nous les avons supprimées pour assurer la consistance de nos résultats.

La littérature dédiée au SEM demande qu'un certain nombre de tests internes soient validés avant d'entamer la lecture et l'interprétation des résultats (Fornell and Larker (1981)). Les tests assurant que nos résultats sont valides sont : *AVE* (pour Average Variance Extracted) - nécessairement au-dessus de 0.5 ; et *FC* (pour Fiabilité Composite) nécessairement au-dessus de 0.7. On peut voir Figure 9 que la plupart de ces tests sont réussis, sauf pour ce qui est de l'*AVE* de Engagement et le *FC* de Challenge et Engagement ; nous pouvons imputer ces résultats un peu faibles à la petite taille de notre échantillon. Cependant, toute proportion gardée, les résultats des tests de validité sont positifs et nous assurent d'avoir des résultats interprétables et pertinents.

**Fiabilité et validité du construit**

Matrice	Alpha de Cronbach	rho_A	Fiabilité composite	Average Variance Extracted (AVE)
	Alpha de Cronbach	rho_A	Fiabilité composite	Average Variance Extra...
CHALLENGE	-0.342	0.939	0.488	0.532
ENGAGEMENT	0.174	0.032	0.283	0.271
IMMERSION	0.439	1.242	0.715	0.587
PERCEPTION D'APPRENTISSAGE	0.554	0.577	0.768	0.527
SKILL_	1.000	1.000	1.000	1.000

Figure 9 – Fiabilité et validité de notre modèle.

## B. Interprétations

Notre modèle corrigé et validé, nous pouvons commencer à interpréter les résultats de cette étude via les tests *PLS* et *Bootstrapping* (voir Figure 10).

Le test *PLS* nous permet d'obtenir des valeurs  $\beta$  (degré de saturation) très intéressantes dans le cadre de cette étude. Nous observons des  $\beta$  très élevés entre Challenge et Perception d'apprentissage ( $\beta = 0.434$ ), Engagement et Perception d'apprentissage ( $\beta = 0.176$ ) et Immersion et Perception d'apprentissage ( $\beta = 0.426$ ). Ce qui signifie que Immersion, Challenge et Engagement sont trois causes explicatives de Perception d'apprentissage. De plus, on remarque que ces variables sont très peu corrélées entre elles : en effet, la saturation entre Engagement et Challenge est de l'ordre de 0.067, et de 0.027 entre Challenge et Immersion. On peut donc en conclure que l'Engagement, le Challenge et la Compétence représentent trois causes explicatives plus ou moins indépendantes de la Perception d'apprentissage. Comme la somme des saturations de ces trois variables est proche de 1, on peut même dire qu'elles expliquent tout à fait la variance de la variable Perception d'apprentissage, ce qui confirme grandement la validité du modèle.

Le *Bootstrapping* permet quant à lui d'obtenir la valeur-t et avec elle le degré de significativité de nos résultats. Nous pouvons donc affirmer l'existence d'une corrélation positive et significative entre Immersion et Perception d'apprentissage et entre Challenge et Perception d'apprentissage. Cela nous permet de confirmer notre hypothèse 1 (H1 : Immersion aura un effet direct positif sur la perception d'apprentissage) ainsi que notre hypothèse 3 (H3 : Challenge aura un effet direct positif sur la perception d'apprentissage.). Même si nous ne pouvons pas confirmer nos autres hypothèses de manière significative – notre faible échantillon en étant certainement la cause –, nos résultats sont très positifs et mettent en exergue l'importance de l'immersion et du challenge dans la conception du jeu éducatif.

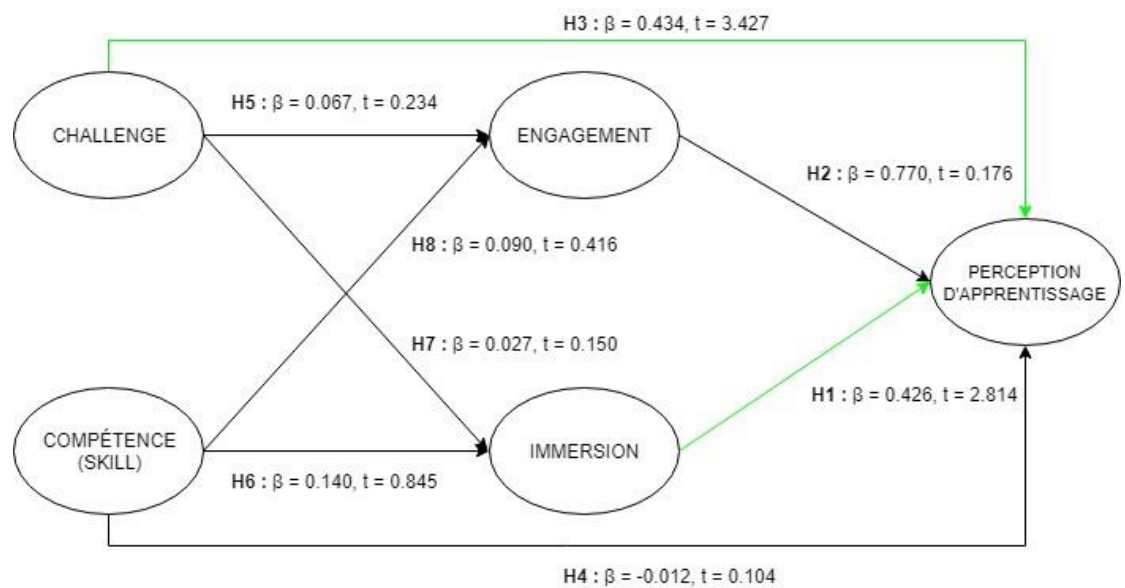


Figure 10 — Résultats des analyses PLS et Bootstrapping. En vert : les corrélations positives significatives (valeur-t significative si supérieure à 1.96).

### 1. Retours qualitatifs - les éléments saillants

Nous avons, en fin d'expérience, posé une série de questions ouvertes (voir Annexe C) aux participants. En plus des éléments notables synthétisés en Annexe D, nous avons constaté que plus l'apprenant s'est senti capable (nombre de level up, ressenti de la difficulté croissante,...), plus il a été immergé et a trouvé le média efficace et utile pour sa prononciation.

Une autre corrélation est également pressentie : l'apprenant qui se déclare relativement bon en anglais (dans le Questionnaire 1) a une perception positive d'apprentissage depuis le média (tout en se déclarant fort immergé). Le tout, indépendamment de son niveau réel et du score qu'il a obtenu en fin d'expérience. Ce qui nous rapproche des mêmes observations que celle de Cheng et al. (2014).

Dans l'ensemble, les retours sur l'expérience sont très positifs. Les participants ont fortement apprécié le jeu et 33 participants sur 52 nous ont demandé d'y avoir accès s'il venait à être disponible publiquement.

La plupart des critiques s'attardaient sur le feedback qui pourrait être plus complet (19 participants sur 52), plus précis et accessible directement durant l'expérience si la phrase prononcée est fautive (afin de mettre en

évidence la faute). Une autre critique concernait les voix des personnages, qui auraient pu sembler plus naturelles (ce qui n'a empêché aucun de nos participant de s'immerger dans le média).

Enfin, une autre piste d'amélioration souvent soulevée est la proposition de plus de choix de réponses et une plus grande variété dans ces dernières (des réponses plus ou moins sérieuses, plus audacieuses, plus humoristiques,...).

## V. Discussion

### A. Comparaison de nos résultats avec la littérature

Si nous ne sommes parvenus à démontrer significativement que deux hypothèses dans notre étude, il est utile de comparer nos résultats avec ceux de la littérature, et spécifiquement avec l'étude qui a inspiré cette recherche.

Hamari et al. (2016) parviennent à démontrer significativement une corrélation positive entre Challenge et Perception d'apprentissage, Engagement et Perception d'apprentissage, ainsi qu'entre le Flow (Challenge et Compétence) et Engagement, et entre le Flow et Immersion. Avec un échantillon composé de 173 individus, Hamari et al. (2016) et leurs deux jeux ne parviennent pas à dénoter une corrélation entre Compétence et Perception d'apprentissage ( $\beta = 0.09$ ,  $t = 0.118$ ), ni entre Immersion et Perception d'apprentissage ( $\beta = 0.084$ ,  $t = 0.785$ ).

Ni notre recherche, ni celle de référence ne parviennent à prouver une corrélation entre Compétence et Perception d'apprentissage. Nos résultats fort similaires représentent Compétence comme la variable la plus faible de notre modèle cherchant à cibler ce qui peut induire la Perception d'apprentissage.

Notre média s'est construit autour de l'immersion en ayant recours à la réalité virtuelle, à un monde réaliste et à une situation narrative plausible. Bien que les jeux 2D de Hamari et al. (2016) ne permettent pas l'immersion comme nous la définissons, ils proposent une mécanique de défi logique et de puzzle beaucoup plus forte que la nôtre.

Cela s'explique par le sujet de l'apprentissage, fort différent du nôtre (dans notre cas : l'amélioration de la prononciation en anglais, dans le leur : l'apprentissage de mécanismes physiques). Il n'est pas surprenant que Challenge chez Hamari et al. (2016) génère des corrélations fortement significatives avec Immersion ( $\beta = 0.728$ ,  $t = 16.878$ ) et avec Perception d'apprentissage ( $\beta = 0.315$ ,  $t = 3.941$ ).

## B. Conclusion

### 1. Résumé

Notre recherche s'est développée en 3 phases :

1. Analyse de la littérature scientifique en *game-based learning* en nous axant sur la question de l'immersion.
2. Réalisation d'un média immersif et interactif en réalité virtuelle servant l'apprentissage ou l'amélioration de la prononciation anglaise.
3. Étude empirique permettant un compte rendu détaillé et une comparaison intéressante avec la littérature scientifique mobilisée.

Nous avons par ailleurs été en mesure de récolter auprès de nos testeurs des pistes d'améliorations du média s'il venait à être continué.

### 2. Limitations

Cette étude se base principalement sur des données auto-reportées par les apprenants. Il serait intéressant de comparer nos résultats avec des données objectives récoltées au sein du média lui-même. Nous avons pu collecter les scores, nombres de level up, nombre de répétitions nécessaires pour chaque apprenant pour aller au bout du média ainsi que la branche narrative qui a été suivie. Nous aurions pu collecter d'autres données utiles comme le temps nécessaire pour réaliser le média ou encore un tracking oculaire complet avec la réalisation de carte de chaleur.

Une deuxième limite est le temps laissé à chaque expérience. Nous aurions voulu analyser des données récoltées suite à des mois d'expérimentations et non quelques minutes. Nous avons développé un épisode de plus ou moins 40 minutes de jeu, et pas des heures de jeu qui peuvent s'étendre sur plusieurs mois d'expérience où questionner l'apprentissage résultant de mois de jeu serait d'autant plus pertinent.

Enfin, nous aurions voulu nous pencher davantage sur la question de la *narration* et y dédier une variable spécifique. Cependant, procéder de cette

manière, en détachant Narration de Immersion, aurait demandé un échantillon plus dense pour permettre l'interprétation des données fiables et pertinentes. La méthodologie propre à SEM demande en effet de respecter ce genre de critères spécifiques et essentiels, garants de la validité de l'étude. Malgré tout, nous avons tenté de constituer une variable Narration composée des questions Skill 2 et Immersion 3 (tenant plus de l'ordre de la *narration* que de la *compétence* et de l'*immersion*). Mais compte tenu de l'échantillon de 52 individus sur lequel repose cette étude, aucun résultat lié à la *narration* n'a pu être établi.

### 3. Contribution de cette étude

Notre recherche examine les conditions de l'immersion, de l'engagement et du flow dans un contexte d'apprentissage. Cette étude prend ses racines dans d'autres recherches questionnant la présence du flow et de l'engagement en contexte de jeu éducatif (par exemple : Hamari et al. (2016) et d'autres qui précèdent : Admiraal et al. (2011), Byun and Loh (2015), Akkerman et al. (2009), Coller and Shernoff (2009), Deater-Deckard et al. (2014), Eseryel et al. (2014), Hou and Li (2014), Hou (2015), Huizenga et al. (2009)). Nos résultats indiquent que l'immersion et le challenge sont deux éléments, dans un contexte ludifié, qui permettent de relever certains défis de l'apprentissage. Et par là même de positionner le jeu vidéo éducatif et immersif comme un moyen potentiellement efficace pour l'amélioration de la prononciation.

## Bibliographie

- Abrantes, S. and Gouveia, L. (2012). Using games for primary school: assessing its use with flow experience. In CruzCunha, M. M., editor, *Handbook of research on serious games as educational, business and research tools*, pages 769–781, Hershey, PA. Information Science Reference.
- Admiraal, W., Huizenga, J., Akkerman, S., and Dam, G. (2011). The concept of flow in collaborative game-based learning. *Computers in Human Behavior*, 27:1185–1194.
- Akkerman, S., Admiraal, W., and Huizenga, J. (2009). Storification in History education: A mobile game in and about medieval Amsterdam. *Computers & Education*, 52:449–459.
- Anderson, J. C. and Gerbing, D. (1988). Structural Equation Modeling in Practice: A Review of Recommended TwoStep Approach. *Psychological Bulletin - PSYCHOL BULL*, 103:411–423.
- Bagozzi, R. and Yi, Y. (2012). Specification, Evaluation, and Interpretation of Structural Equation Models. *Journal of the Academy of Marketing Science*, 40:8–34.
- Brougère, G. (2012). Le jeu peut-il être sérieux ? Revisiter ~ Jouer/Apprendre en temps de serious game. *Australian Journal of French Studies*, 49:117–129.
- Byun, J. and Loh, C. (2015). Audial engagement: Effects of game sound on learner engagement in digital game-based learning environments. *Computers in Human Behavior*.
- Caïra, O. (2018). Les dimensions multiples de l'engagement ludique. *Sciences du jeu*.
- Cheng, M.-T., She, H.-C., and Annetta, L. (2014). Game immersion experience: Its hierarchical structure and impact on game-based science learning. *Journal of Computer Assisted Learning*, 31.
- Chin, W., Marcolin, B. L., and Newsted, P. R. (2003). A Partial Least Squares Latent Variable Modeling Approach for Measuring Interaction Effects: Results from a Monte Carlo Simulation Study and

- an Electronic-Mail Emotion/Adoption Study. *Information Systems Research*, 14:189–217.
- Coller, B. and Shernoff, D. (2009). Video Game-Based Education in Mechanical Engineering: A Look at Student Engagement. *International Journal of Engineering Education*, 25.
  - Csikszentmihalyi, M. (1991). *Flow: The Psychology of Optimal Experience*, volume 8.
  - Csikszentmihalyi, M. (1997). *Creativity: Flow and the Psychology of Discovery and Invention*.
  - Csikszentmihalyi, M. (2008). *Flow, The Psychology of Optimal Experience*. New York. Harper Perennial Modern Classics.
  - Csikszentmihalyi, M., Rathunde, K., and Whalen, S. (1993). *Talented Teenagers: The Roots of Success and Failure*, volume 42.
  - Davidson, D. (2008). *Beyond fun: Serious games and media*. ETC Press, Pittsburgh, PA.
  - Deater-Deckard, K., Mallah, S. E., Chang, M., Evans, M., and Norton, A. (2014). Student behavioral engagement during mathematics educational video game instruction with 11–14 year olds. *International Journal of Child-Computer Interaction*, 2.
  - Ermi, L. and Mäyrä, F. (2005). Fundamental Components of the Gameplay Experience: Analysing Immersion. In *Worlds in Play: Int. Perspectives on Digital Games Research*.
  - Eseryel, D., Law, V., Ifenthaler, D., Ge, X., and Miller, R. (2014). An Investigation of the Interrelationships between Motivation, Engagement, and Complex Problem Solving in Game-based Learning. *Educational Technology & Society*, 17:42–53.
  - Fassbender, E., Richards, D., Bilgin, A., Thompson, W., and Heiden, W. (2012). VirSchool: The effect of background music and immersive display systems on memory for facts learned in an educational virtual environment. *Computers & Education*, 58:490–500.
  - Fornell, C. and Larcker, D. F. (1981). Evaluating Structural Equation Models With Unobservable Variable and Measurement Error. *Journal of Marketing Research*, 18:39–50.

- Hair, J., Black, W., Babin, B., and Anderson, R. (2009). *Multivariate data analysis*. Prentice Hall. *London*.
- Hair, J. F. J., Anderson, R., Tatham, R. L., and Black, W. (1998). *Multivariate Data Analysis, 5th Ed.*
- Hamari, J. and Koivisto, J. (2015). Why do people use gamification services? *International Journal of Information Management*, 35:419–431.
- Hamari, J., Shernoff, D., Rowe, E., Coller, B., Asbell-Clarke, J., and Edwards, T. (2016). Challenging games help students learn: An empirical study on engagement, flow and immersion in gamebased learning. *Computers in Human Behavior*.
- Hou, H.-T. (2015). Integrating cluster and sequential analysis to explore learners' flow and behavioral patterns in a simulation game with situated-learning context for science courses: A video-based process exploration. *Computers in Human Behavior*, 48.
- Hou, H.-T. and Li, M.-C. (2014). Evaluating multiple aspects of a digital educational problem-solving-based adventure game. *Computers in Human Behavior*, 30:29–38.
- Huizenga, J., Admiraal, W., Akkerman, S., and Dam, G. (2009). Mobile game-based learning in secondary education: Engagement, motivation and learning in a mobile city game. *J. Comp. Assisted Learning*, 25:332–344.
- Joyner, K., Csikszentmihalyi, M., and Schneider, B. (2002). Becoming Adult: How Teenagers Prepare for the World of Work. *Contemporary Sociology*, 31:140.
- Nicholls, J. G. (1979). Quality and equality in intellectual development: The role of motivation in education. *American Psychologist*, 34:1071–1084.
- Ringle, C. M., Wende, S., and Will, A. (2005). *SmartPLS*, 2(0):3.
- Thomas, J. W. (1980). Agency and Achievement: Self-management and Self-regard. *Review of Educational Research - REV EDUC RES*, 50:213–240.
- White, R. W. (1959). Motivation Reconsidered: The Concept Of Competence. *Psychological review*, 66:297–333.

## Annexe A. Questionnaire 1 démographique

Ce questionnaire sert à cerner le niveau d'anglais (général et prononciation) auto déclaré par l'utilisateur et ses habitudes d'utilisation de la VR, en plus des informations démographiques communément demandées.

Q1.1 Quel est votre genre : Femme – Homme – Autre

Q1.2 Quel âge avez-vous ?

Q1.3 Quelle est votre langue maternelle ?

Q1.4 Autres langues parlées ?

Q1.5 Comment estimez-vous votre niveau d'anglais général (connaissance du vocabulaire, de la grammaire, de la conjugaison...) ? (Echelle de Likert de 1 à 5.)

Q1.6 Comment estimez-vous votre niveau de prononciation en anglais ? (Echelle de Likert de 1 à 5.)

Q1.7 A quelle fréquence utilisez-vous l'anglais oralement ? (Tous les jours / 2 à 3x par semaine / 1x par semaine / Rarement (moins d'une fois par mois) / Jamais ou presque jamais.)

Q1.8 Dans quel contexte avez-vous recours à l'anglais oralement ? (situation scolaire, professionnelle, familiale, autre)

Q1.9 Quel est votre usage d'outils technologiques servant l'apprentissage des langues ? (Echelle de Likert de 1 à 5.)

Q1.10 Quel est votre usage de la VR ? (Echelle de Likert de 1 à 5.) Lien vers le questionnaire en ligne : <https://graal.typeform.com/to/oQCG00>

## Annexe B. Questionnaire 2 - retours d'expérience

Q2.1 (Apprentissage 1) Pensez-vous pouvoir améliorer votre prononciation en utilisant quotidiennement ce média ?

Q2.2 (Apprentissage 2) Avez-vous eu l'impression d'entraîner votre prononciation ?

Q2.3 (Apprentissage 3) Le média vous aide-t-il à progresser dans votre prononciation ?

Q2.4 (Engagement 1) A quel point étiez-vous concentré ?

Q2.5 (Engagement 2) A quel point avez-vous apprécié ce que vous faisiez ?

Q2.6 (Engagement 3) A quel point était-ce divertissant d'interagir avec les personnages ?

Q2.7 (Engagement 4) A quel point le média était-il intéressant ?

Q2.8 (Engagement 5 – reverse) A quel point vous êtes-vous ennuyé ?

Q2.9 (Immersion 1) Avez-vous perdu la notion du temps en utilisant le média ?

Q2.10 (Immersion 2) A quel point étiez-vous immergé dans le média ?

Q2.11 (Immersion 3) A quel point désiriez-vous découvrir à qui appartenaient les places de concert égarées ?

Q2.12 (Challenge 1) Le média vous a-t-il challengé ?

Q2.13 (Challenge 2) Le média vous pousse-t-il aux limites de votre niveau ?

Q2.14 (Challenge 3) Avez-vous trouvé le média trop facile par rapport à votre niveau de prononciation ?

Q2.15 (Skill 1) A quel point étiez-vous bon dans la prononciation de vos phrases dans le média ?

Q2.16 (Skill 2) Avez-vous choisi le plus souvent les phrases qui vous semblaient les plus difficiles parmi les propositions ?

## Annexe C. Questionnaire 3 - Questions qualitatives : avis et améliorations

Q3.1 Qu'avez-vous pensé de votre expérience ?

Q3.2 Quel personnage avez-vous préféré et pourquoi ?

Q3.3 Avez-vous remarqué que Markus était aveugle ? (Si oui, comment ?)

Q3.4 Qu'avez-vous pensé de la qualité des graphismes ?

Q3.5 Qu'amélioriez-vous dans le média dans son ensemble ?

Q3.6 Qu'avez-vous pensé de la voix des personnages ?

Q3.7 Avez-vous senti que vous étiez plus ou moins mis en difficulté au long  
de votre expérience ? (Si oui, quand ?)

Q3.8 Auriez-vous envie de refaire l'expérience ou de rejouer au jeu ?

## Annexe D. Synthèse des commentaires du questionnaire 3



Voici un schéma reprenant toutes les idées-clés qui ont été formulées lors des questions ouvertes. Taille et ordre sont relatifs au degré de récurrence des avis convergents (les nombres sur le schéma reprennent les participants ayant exprimé l'idée) sur un total de 52 participants.

## Table des matières

I.	Introduction .....	5
A.	Contexte et démarche .....	5
B.	Définitions .....	5
1.	"Serious" game et gamification.....	5
2.	Concept de flow .....	6
3.	Concept d'immersion.....	8
4.	La perception d'apprentissage .....	9
C.	Article de référence.....	10
II.	Média.....	12
A.	Description du média.....	12
1.	Tutoriel — 1 <sup>ère</sup> partie.....	12
2.	L'épisode du train — 2 <sup>ème</sup> partie .....	13
3.	Feedback et score — 3 <sup>ème</sup> partie.....	17
B.	Conception du média .....	18
1.	Contexte .....	18
2.	Réalisation.....	19
C.	Limites et difficultés .....	26
III.	Méthode.....	30
A.	Participants et procédures .....	30
B.	Validité .....	31
IV.	Résultats.....	33
A.	Corrections et adaptations du modèle.....	33
B.	Interprétations .....	35
1.	Retours qualitatifs - les éléments saillants .....	36
V.	Discussion .....	38

A. Comparaison de nos résultats avec la littérature .....	38
B. Conclusion .....	39
1. Résumé.....	39
2. Limitations .....	39
3. Contribution de cette étude .....	40
Bibliographie.....	41
Annexe A. Questionnaire 1 démographique .....	44
Annexe B. Questionnaire 2 - retours d'expérience .....	45
Annexe C. Questionnaire 3 - Questions qualitatives : avis et améliorations .....	46
Annexe D. Synthèse des commentaires du questionnaire 3.....	47

Nous investiguons dans cette étude l'influence de l'immersion, du *flow* et de l'engagement sur l'apprentissage (ou l'amélioration) de l'accent anglais au travers d'un média en réalité virtuelle. Les données ont été récoltées sur un échantillon de 52 testeurs directement après qu'ils aient découvert et utilisé le média en réalité virtuelle. Les résultats montrent que l'immersion et le challenge ont un effet positif clair sur l'apprentissage. Cependant, nous n'avons pas trouvé d'effet significatif entre l'engagement et l'apprentissage ni entre les compétences (Skills) et l'apprentissage. Dans cette étude alliant réalité virtuelle et apprentissage, l'immersion fut l'élément mis en exergue pour en observer les effets dans un contexte éducatif. L'immersion et le challenge sont de bons prédicteurs de l'apprentissage : nos résultats suggèrent qu'ils peuvent assurer aux apprenants de persévérer dans leur apprentissage.

Immersion – Apprentissage par le jeu – Réalité virtuelle – Flow – Prononciation de l'anglais – Serious game.