Université de Montréal

Caractérisation du niveau d’amusement grâce à des techniques d’apprentissage machine

Par

Gabrielle Toupin

Département de psychologie, Faculté des arts et des sciences

Mémoire présenté en vue de l’obtention du grade de

Maîtrise ès sciences (M. Sc.) en psychologie

Mai 2021

Université de Montréal

Département de psychologie, Faculté des arts et des sciences

Ce mémoire intitulé

Caractérisation du niveau d'amusement grâce à des techniques d'apprentissage machine

Présenté par

Gabrielle Toupin

A été évalué(e) par un jury composé des personnes suivantes

Pierre Bellec
Président-rapporteur

Karim Jerbi
Directeur de recherche

Claude Frasson
Codirecteur

Sébastien Hétu
Membre du jury

Résumé

Introduction. L'humour est un processus cognitif complexe qui peut entraîner un état émotionnel positif d'amusement. La réponse émotionnelle déclenchée par l'humour possède plusieurs bénéfices pour la santé. Son utilisation en recherche et lors d'essais cliniques est d'ailleurs de plus en plus fréquente. Malheureusement, l'appréciation de l'humour varie considérablement d'un individu à l'autre, et entraîne des réponses émotionnelles très différentes. Cette variabilité, rarement prise en compte dans les études de recherche, est donc importante à quantifier pour pouvoir évaluer de manière robuste les effets de l'humour sur la santé. **Objectifs.** Ce projet de maîtrise vise à explorer différentes modalités permettant d'établir une mesure objective de l'appréciation de l'humour via des techniques d'apprentissage automatique et d'apprentissage profond. Les caractéristiques de la vidéo, les expressions faciales et l'activité cérébrale ont été testées comme prédicteur potentiels de l'intensité de l'amusement. **Étude 1.** Dans notre première étude, les participants (n = 40) ont regardé et évalué des vidéos humoristiques et neutres pendant que leurs expressions faciales étaient enregistrées. Pour chaque vidéo, nous avons calculé le mouvement moyen, la saillance et deux scores sémantiques. L'algorithme d'arbres aléatoire a été entraîné sur les caractéristiques des vidéos et le sourire des participants afin de prédire à quel point le participant a évalué la vidéo comme étant drôle, et ce, à trois moments durant la vidéo (début, milieu et fin). De plus, nous avons utilisé l'expression faciale du participant pour explorer la dynamique temporelle de l'appréciation de l'humour tout au long de la vidéo et ses impacts sur la vidéo suivante. Nos résultats ont montré que les caractéristiques des vidéos permettent de bien classifier les vidéos neutres et les vidéos humoristiques, mais ne permettent pas de différencier les intensités d'humour. À l'inverse, le sourire est un bon prédicteur de l'intensité de l'amusement au sein des vidéos humoristiques (contribution=0.53) et est la seule modalité à fluctuer dans le temps; montrant ainsi que l'appréciation de l'humour est plus grande à la fin de la vidéo et après la vidéo. **Étude 2.** Notre deuxième étude a utilisé des techniques d'apprentissage profond afin de prédire l'intensité de l'amusement ressenti par les participants (n = 10) lorsqu'ils visionnaient des vidéos humoristiques avec un casque EEG commercial. Nous avons utilisé un algorithme LSTM pour prédire les intensités d'amusement (faible, modéré, élevé, très élevé) en fonction d'une seconde d'activité cérébrale. Les résultats ont montré une bonne transférabilité entre les participants et une précision de décodage dépassant 80% d'exactitude. **Conclusion.** Les caractéristiques de la vidéo, les expressions faciales des participants et l'activité cérébrale ont permis de prédire l'appréciation de l'humour. À partir de ces trois modalités, nous avons trouvé que les réactions physiologiques (expression faciale et activité cérébrale) prédisent mieux les intensités de l'amusement tout en offrant une meilleure précision temporelle de la dynamique d'appréciation de l'humour. Les futures études employant l'humour gagneraient à inclure le niveau d'appréciation, mesuré via le sourire ou l'activité cérébrale, comme variable d'intérêt dans leurs protocoles expérimentaux.

Mots-clés : amusement, humour, apprentissage machine, apprentissage profond, LSTM, Forêt d'arbres décisionnels

Abstract

Introduction. Humour is a complex cognitive process that can result in a positive emotional state of amusement. The emotional response triggered by humour has several health benefits and is used in many research and clinical trials as treatments. Humour appreciation varies greatly between participants and can trigger different levels of emotional response. Unfortunately, research rarely considers these individual differences, which could impact the implication of humour in research. These researches would benefit from having an objective method to detect humour appreciation. **Objectives.** This master's thesis seeks to provide an appropriate solution for an objective measure of humour appreciation by using machine learning and deep learning techniques to predict how individuals react to humorous videos. Video characteristics, facial expressions and brain activity were tested as potential predictors of amusement's intensity. **Study 1.** In our first study, participants (n=40) watched and rated humorous and neutral videos while their facial expressions were recorded. For each video, we computed the average movement, saliency and semantics associated with the video. Random Forest Classifier was used to predict how funny the participant rated the video at three moments during the clip (begging, middle, end) based on the video's characteristics and the smiles of the participant. Furthermore, we used the participant's facial expression to explore the temporal dynamics of humour appreciation throughout the video and its impacts on the following video. Our results showed that video characteristics are better to classify between neutral and humorous videos but cannot differentiate humour intensities. On the other hand, smiling was better to determine how funny the humorous videos were rated. The proportion of smiles also had more significant fluctuations in time, showing that humour appreciation is greater at the end of the video and the moment just after. **Study 2.** Our second study used deep learning techniques to predict how funny participants (n=10) rated humorous videos with a commercial EEG headset. We used an LSTM algorithm to predict the intensities of amusement (low, medium, high, very high) based on one second of brain activity. Results showed good transferability across participants, and decoding accuracy reached over 80%. **Conclusion.** Video characteristics, participant's facial expressions and brain activity allowed us to predict humour appreciation. From these three, we found that physiological reactions (facial expression and brain activity) better predict funniness intensities while also offering a better temporal precision as to when humour appreciation occurs. Further studies using humour would benefit from adding physiological responses as a variable of interest in their experimental protocol.

Keywords: amusement, humour, machine learning, deep learning, LSTM, Random Forest

Table des matières

Liste des tableaux

Liste des figures

Liste des sigles et abréviations

ANOVA : Analyse de la variance

AUs : Action Units

AU12 : Muscle Zygomatique Majeur (Action Unit)

AU6 : Muscle Orbiculaire de l'œil (Action Unit)

EEG : Électroencéphalographie

Hz: Hertz

ICA: Independent Component Analysis

LSTM: Long-Short Term Memory

NGD: Normalized Google Distance

RFC: Random Forest Classifier

SHQ-6 : Sense of Humour Questionnaire

Remerciements

J’aimerais tout d’abord remercier mon directeur de recherche Karim Jerbi pour m’avoir accepté dans son laboratoire. Mon passage dans le Coco Lab m’a permis d’apprendre énormément sur la programmation et l’intelligence artificielle, deux domaines qui me tiennent énormément à cœur et qui sont maintenant centrales à mes projets futurs. Je tiens également à remercier Anne-Lise Saive pour m’avoir guidé au travers de ce projet grâce à ses conseils et ses nombreuses idées, et qui m’a aidé à adapter le projet en cours de route lorsque les passations en présentiels n’étaient plus possibles.

J’aimerais également remercier Yan Cyr pour m’avoir offert l’opportunité de conduire ma maîtrise en partenariat avec l’entreprise BMU et qui a financé ma maîtrise, le professeur Claude Frasson pour avoir supervisé ce projet appliqué ainsi que Sahbi Benlamine pour avoir aidé avec le développement et l’implémentation de l’outil de détection d’humour.

Finalement, je tiens à remercier mon conjoint et mes parents pour leur soutien moral et leurs encouragements tout au long de ce projet. Notamment, j’aimerais souligner le support continuel de mon conjoint Esteban qui m’a poussé à me surpasser, à faire face aux défis et qui m’a encouragée à continuer même lorsque ça n’allait pas comme prévu.

Merci à tous ceux qui ont contribué à ce projet.

Chapitre 1 – Introduction

Qu’est-ce que l’humour?

Théorie de la Résolution d'incongruité

Que ce soit sous forme de divertissement, tel que de visionner un film (ex: comédie, parodie), ou lors de nos interactions sociales (ex: dire une blague à un collègue), l’humour fait partie intégrante de notre quotidien. L’humour peut être suscité de manière auditive (ex: blague) ou visuelle (ex: image) et peut comprendre un aspect langagier ou non. Quelle que soit la manière dont le contenu humoristique est présenté, les recherches s’entendent pour dire que l’humour est composé d’un processus cognitif et d’un processus émotionnel (Vrticka et al., 2013). Selon la théorie de la résolution d’incongruité (Suls, 1972), le processus cognitif de l’humour se fait en deux parties: la détection d’incongruités et la résolution.

L’humour est un processus cognitif complexe qui requiert une mise en situation, soit le contexte dans lequel la situation humoristique se produit. Elle correspond généralement à une situation où la fin est évidente et attendue. La mise en situation est suivie d’un élément inattendu, tel que la réponse à une punchline ou la suite de l’événement, qui déclenche l’humour. Le processus cognitif associé à l’humour est premièrement composé de la détection d’un élément incongru et inattendu. Le processus cognitif se complète ensuite par la résolution de l’incongruité, faisant appel à nos souvenirs et à nos connaissances, nous réinterprétons comment la mise en situation a pu se conclure de cette manière. La détection et la résolution des incongruités génèrent ainsi l’amusement lié à l’humour.

Un exemple d’humour visuel peut être trouvé dans la figure 1. La mise en situation se situe dans la vignette **A** où on peut y voir une personne se noyer et un homme qui s’inquiète. La suite logique et attendue serait une scène de sauvetage. La suite humoristique se trouve à la vignette **B**, où on aperçoit que l’homme vole la montre de la personne dans l’eau. Cette fin diffère de la situation attendue qui devrait être un sauvetage et résulte en une situation humoristique.

L'appréciation de l'humour est caractérisée par le niveau d'amusement généré. L'état émotionnel évoqué par l'humour est décrit de manière très variée dans la littérature. Encore aujourd'hui, on y associe des états émotionnels tels que de la joie, l'amusement ou encore de l'excitation (Martin et Ford, 2018). Un contenu humoristique drôle engendrera souvent, mais pas toujours, un sourire ou un rire. À des fins de simplicité, le terme *amusement* référé tout au long de ce mémoire fait référence aux émotions positives induites par le contenu humoristique.

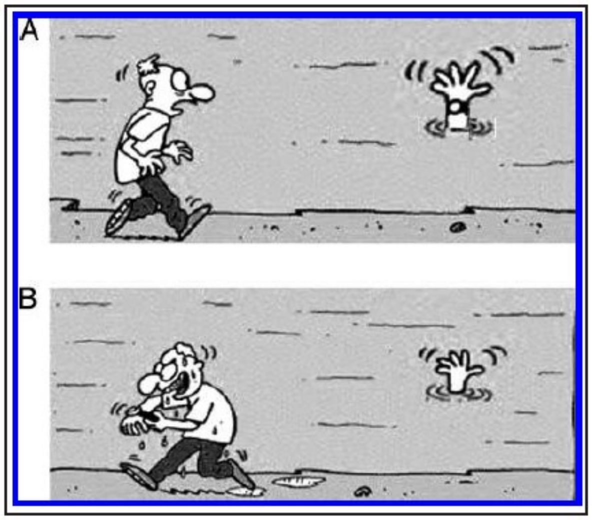

Figure 1. – **Exemple d'humour visuel.** Représentation visuelle du processus cognitif associé à l'humour. **(A)** correspond à la mise en situation et **(B)** correspond à une fin humoristique. L'alternative non-humoristique et attendue serait une case où l'homme sauve la personne qui se noie. Bande dessinée prise de l'étude de Bartolo et al. (2016)

Bienfaits et avantages de l'humour

Que ce soit via des interactions sociales (ex: écouter la blague d'un collègue) ou le visionnement de contenus humoristiques (ex: film, écouter la reprise vidéo d'un comédien, etc.), l'humour a un impact direct sur les relations sociales et la santé.

L'humour est un comportement social qui peut être généré de manière volontaire ou non. On retrouve cette interaction dans de nombreuses sphères de vie où les interactions sociales sont centrales, notamment dans les milieux de travail (Mathies et al., 2016). En effet, l'humour est utilisé par les travailleurs de la santé afin de réduire la tension et le stress, améliorer la collaboration avec les patients et aider à la gestion d'émotions (Ghaffari et al., 2015; Dean et

Major, 2008). L'humour est également utilisé dans les milieux de travail où il facilite le travail d'équipe en augmentant la cohésion du groupe et la communication, tout en offrant un meilleur leadership et une meilleure culture d'entreprise (Mathew et Vijayalakshmi, 2017; Gozukara, 2016; Murata, 2014). L'humour a aussi des avantages dans le milieu scolaire et de l'enseignement. L'utilisation d'humour par les professeurs permet de rendre les élèves plus confortables avec l'enseignant et d'améliorer l'attention, la concentration et l'intérêt des étudiants, favorisant ainsi un meilleur environnement d'apprentissage (Baid et Lambert, 2010; Shatz et LoSchiavo, 2006).

En plus de posséder des bienfaits dans les interactions sociales et les sphères de vie associées, l'intégration de l'humour dans le traitement de maladies est de plus en plus explorée. La thérapie par humour contient peu de risque, est très accessible et peu coûteuse, rendant son utilisation très attrayante (Zhao et al., 2019; Gonot-Schoupinsky et Garip, 2018; Berk et al., 2001). La thérapie par l'humour consiste généralement en une ou plusieurs périodes où du contenu humoristique, tel que de courts extraits cinématiques, des spectacles de comédiens ou même des clowns qui racontent des blagues, sont présentés. Plusieurs composantes cognitives (ex: mémoire) et émotionnelles (ex: état anxieux) sont comparées avant et après la thérapie.

La thérapie par humour est utilisée autant dans les problèmes de santé physique (ex: crise cardiaque), qu'au niveau de la santé mentale (ex: dépression) et des processus cognitifs (ex: mémoire). L'humour influence différents systèmes physiologiques, tels que la réduction des hormones de stress (ex: cortisol), permettant ainsi de réduire l'anxiété et certains effets de la dépression à court et long terme (Zhao et al., 2019; Colom et al., 2011; Ko et Youn, 2011; Fonzi, et al. 2010; Berk et al., 2001; Berk et al., 1989). On observe également une augmentation du bien-être général et des sentiments comme l'espoir et l'optimisme à la suite du visionnement de vidéos humoristiques (Zhao et al, 2020; Crawford et Caltabiano, 2011; Vilaythong et al., 2003). L'humour, tel que les comédies, peut également agir comme distracteur afin de réduire la douleur. Les gens ayant un plus grand sens de l'humour sont généralement plus résilients à la douleur et aux stresseurs quotidiens (Pérez-Aranda et al., 2019; Behrouz et al., 2017). Sur un plan plus physique, des séances humoristiques aident à réduire les crises cardiaques récurrentes (Tan et al., 2007) et les douleurs chroniques (Behrouz et al, 2017).

Des goûts humoristiques différents, une problématique?

Les études utilisant l'humour comme thérapie sont très prometteuses. Toutefois, à quel point le contenu est perçu comme drôle peut affecter l'efficacité de la thérapie (Froehlich et al., 2021; Moran et al, 1996). Récemment, Froehlich et ses collègues (2021) ont montré que le stress ressenti et le niveau de cortisol dans la salive des participants étaient beaucoup moindre chez ceux ayant préalablement visionné un film humoristique contrairement à leurs pairs ayant visionné une vidéo neutre. Mais plus important encore, leur étude montre que les participants qui évaluaient le clip humoristique comme étant plus drôle que leurs pairs arrivaient à supporter un niveau de douleur plus élevé, montrant ainsi que le niveau d'appréciation de l'humour a un impact physiologique qui est plus que binaire.

Un même contenu humoristique peut être perçu comme excessivement drôle par une personne mais de manière médiocre par une autre. Pourtant, la majorité des études, comme celle de Froehlich (2021), utilisent un stimulus unique (ex: le même extrait de film est vu par tous les participants) pour générer l'amusement. Bien que le contenu soit généralement présélectionné spécifiquement pour induire de l'amusement, la nature très subjective de l'humour rend son utilisation dans des procédures standardisées compliquée. Par exemple: le *participant A* peut avoir trouvé toutes les blagues très drôles tandis que le participant *B* n'a trouvé qu'une seule blague drôle. En plus d'utiliser un stimulus unique, très peu d'études prennent la peine de mesurer la perception du participant vis-à-vis ce contenu.

Il serait donc important de pouvoir mesurer de manière objective le sentiment d'amusement. En effet, une méthode objective permettrait une meilleure modulation des résultats lors de futures études utilisant l'humour. Au vue de la grande variabilité interindividuelle du sens de l'humour, une mesure d'appréciation de l'humour (*à quel point le participant à trouver le contenu drôle*) objective et automatique permettrait de caractériser précisément l'effet de l'humour dans les protocoles de recherche lorsque l'évaluation subjective du participant n'est pas possible. Une technique objective de mesure de l'humour serait favorable, par exemple, lorsque le stimulus humoristique (ex : clown qui raconte des blagues) et la tâche (ex : tâche cognitive) sont performé sur une période prolongée et conduite simultanément.

Ainsi, le présent projet de mémoire vise à explorer différentes techniques qui permettraient de quantifier objectivement l'état d'amusement à la suite d'un stimulus humoristique. Les sections suivantes de l'introduction décrivent en plus de détails **1)** les différences individuelles connues qui se retrouvent dans l'appréciation de l'humour, **2)** les échelles comportementales actuellement utilisées dans le domaine pour mesurer l'appréciation humoristique ainsi que **3)** différentes pistes qui permettraient d'évaluer objectivement le niveau d'amusement.

Variabilité individuelle dans l'humour

Les réactions associées à l'humour varient énormément. Cette variabilité se retrouve à différentes étapes du processus humoristique. En effet, plusieurs caractéristiques individuelles influencent la production (ex: *quel type de blague on raconte*), l'appréciation (ex: *quel type de blague on préfère écouter*) et la réponse émotionnelle (ex: *à quel point nous rions*). Ces différences individuelles se retrouvent autant au niveau du genre (Wu et al., 2016; Bergen, 2009), de l'âge (Henry et al., 2013; Svebak et al., 2004) et de l'éducation (Bischetti et al., 2021; Willinger et al., 2017) qu'au niveau de l'état psychologique actuel (ex: *humeur*) et physique (Ehrlé et al., 2020; Thaler et al., 2012).

Influence démographique

Afin de mieux comprendre les différences individuelles quant à la production et l'appréciation de l'humour, les recherches se sont intéressées aux caractéristiques démographiques des gens. Plus précisément, les recherches ont avidement exploré les différences entre les hommes et les femmes. La différence la plus fréquente se retrouve lors de la production et l'appréciation de l'humour agressif où les hommes montrent une plus grande préférence que les femmes (Wu et al., 2016; Bergen, 2009). Toutefois, le niveau d'empathie peut moduler l'effet du genre sur l'humour agressif. Ainsi, une plus grande tendance à utiliser un humour agressif est souvent associée à une difficulté à percevoir la souffrance d'autrui (Wu et al., 2016). Les hommes ont également tendance à considérer l'humour sexuel et l'humour noire comme étant plus drôle que les femmes (Carretero-Dios et Ruch, 2010; Barrick et al., 1990). L'expression des émotions et du plaisir vis-à-vis de l’humour serait en revanche plus présente chez les femmes que chez les hommes (Brody, 2000). Cette différence pourrait être due à un usage automatique de la régulation des émotions chez les hommes, ce qui réduirait l’activité de régions cérébrales clés dans l’appréciation et l’expression de l’humour, comme l'amygdale, les aires préfrontales et le striatum ventral (Kohn et al., 2011; McRae et al., 2008).

L'âge est également une caractéristique distinctive de l'humour. Le sens de l'humour a tendance à décliner avec l'âge. De plus, les personnes âgées ont plus de difficultés à comprendre l'humour et ils expriment moins leur amusement (Henry et al., 2013; Svebak et al., 2004). Cette différence se retrouve principalement chez les personnes de plus de 65 ans et serait associée à une détérioration du cortex frontal (Shammi et Stuss, 2003).

Plus récemment, les recherches se sont intéressées à l’implication du niveau d’éducation dans la production et l’appréciation de l’humour. Notamment, les recherches ont montré que les personnes moins scolarisées ont tendance à juger l'humour comme étant plus drôle, tandis que les personnes ayant suivi une éducation supérieure sont plus susceptibles de comprendre et de préférer l'humour noir (Bischetti et al., 2021; Willinger et al., 2017).

Santé physique et psychologique

La santé physique et psychologique d’un individu est aussi liée à son sens de l’humour. Notamment, les personnes indiquant qu’ils sont globalement satisfaites de leur santé ont tendance à avoir un sens de l'humour plus élevé (Svebak et al., 2004).

Une santé psychologique ou physique moindre est souvent associée à un humour moindre. En effet, les personnes atteintes de troubles mentaux, telles que la dépression, ont tendance à avoir un sens de l'humour plus faible (Svebak, 2010; Thorson et al, 1997). Cette faiblesse pourrait être expliquée par une difficulté cognitive à attribuer des états mentaux inobservable aux protagonistes dans la blague (Uekermann et al., 2008). Dans la même direction, les personnes atteintes de troubles neurologiques et neurodégénératifs ont tendance à considérer l'humour comme moins drôle et ont également un sens de l'humour plus faible (Ehrlé et al., 2020; Thaler et al., 2012).

Mesurer les différences individuelles : échelles comportementales

Plusieurs chercheurs en psychologie ont créé des échelles comportementales afin d'évaluer les différences inter-individuelles associées à la production et l'appréciation d'humour. L'utilisation de ces échelles avant une tâche pourrait permettre de mieux comprendre quels participants évalueraient les vidéos comme étant plus drôles sans avoir à évaluer chaque stimulus individuellement.

Les deux principaux questionnaires utilisés dans la littérature sont le questionnaire sur le style d'humour (*Humour Style Questionnaire*; HSQ) de Martin et al. (2003) et le questionnaire sur le sens de l'humour (*sense of humour questionnaire;* SHQ-6) de Svebak (2010, 1996).

Échelles des styles humoristiques

La théorie de Martin et al. (2003) décrit, via deux dimensions, quatre raisons d'utiliser l'humour. Les deux dimensions sont 1) l'adaptation (adaptée ou mésadaptée) et 2) la cible de l'humour (en lien avec les relations sociales ou envers soi-même). Du côté adapté on retrouve l'humour *affiliatif* qui correspond à un type d'humour tourné vers autrui où l'humour est utilisé afin de faire rire, de renforcer les liens sociaux et de réduire la tension sociale. L'humour dit d'*auto-défense* est quant à lui tourné vers soi-même et aide à faire face aux événements stressants et permet de voir le bon côté d'une situation. L'humour affiliatif et d'auto-défense sont tous deux corrélés avec le sens de l'humour de Svebak (1996) (Martin et al.,2003). Du côté mésadapté de l'humour, on retrouve un humour tourné vers autrui: l'humour *agressif*. Ce type d'humour est utilisé au détriment d'autrui et l'utilisation de sarcasme et de critiques est utilisée dans le but de ridiculiser ou rabaisser la personne visée. Finalement, le dernier type *auto-dérisoire* consiste en un humour où l'on fait rire les autres au détriment de soi-même dans le but de se faire valoir.

Bien que l'échelle de Martin (2003) pourrait être utilisée pour anticiper le niveau d'appréciation humoristique perçu par un participant, il faudrait également prendre en considération le contenu du média visionné (ex : niveau d'agressivité de la vidéo). Ce dernier rend la prédiction beaucoup plus complexe.

Questionnaire sur le sens de l'humour

L'échelle de Svebak (2010, 1996) vise à définir selon trois dimensions la tendance de l'individu à percevoir, apprécier et utiliser l'humour dans la vie de tous les jours. On observe une première dimension *cognitive* qui détermine l'habileté de l'individu à utiliser et percevoir le caractère humoristique d'une situation. La dimension *sociale* correspond davantage à l'appréciation de l'humour dans les relations avec autrui et la dimension *affective* correspond à la tendance à réagir positivement (ex: rire) à l'humour. La combinaison des trois dimensions définit l'appréciation générale de l'humour dans la vie de tous les jours.

Avec ses six questions, la version courte de l'échelle mesurant le sens de l'humour (SHQ-6), tel que décrit par Svebak (2010), pourrait être intégré facilement dans les recherches employant l'humour afin de modérer les résultats. Toutefois, il s'agirait d'un indice général et non-spécifique à un stimulus donné. Il est donc nécessaire d'approfondir différentes techniques qui permettraient une mesure objective d'un stimulus humoristique.

Prédiction objective de l'appréciation humoristique

Le développement d'outils permettant d'estimer de manière objective le niveau d'appréciation de l'humour serait bénéfique aux recherches et aux essais cliniques utilisant l'humour. Wang et Ji (2015) ont récemment recensé les différentes techniques objectives les plus couramment utilisées permettant d'évaluer le contenu affectif et la réaction du participant lors de visionnement de courtes vidéos. Ils divisent les différents moyens de prédire l'état émotionnel en deux catégories: direct et implicite. Une approche directe utilise le contenu et les caractéristiques audiovisuelles de la vidéo (ex: intensité du mouvement dans la scène), tandis que l'évaluation implicite se base sur les réponses physiologiques du participant (ex: expressions faciales).

Trois avenues, détaillées plus en profondeur ci-bas, semblent les plus prometteuses afin de prédire l'amusement: les caractéristiques de la vidéo, le sourire et les mesures physiologiques telles que l'électroencéphalographie (EEG).

Prédiction grâce aux caractéristiques de la vidéo

L'utilisation des caractéristiques et du contenu d'une vidéo pour prédire à quel point la vidéo est drôle peut être envisagée. Les réalisateurs de films utilisent souvent les caractéristiques de la vidéo, telles que le mouvement, pour induire des émotions différentes et ces mêmes caractéristiques peuvent, en retour, permettre de prédire le genre du film (ex: action, romance, horreur) (Rasheed et al., 2005). Utilisant des techniques d'apprentissage machine (voir encadré 1), plusieurs auteurs arrivent à prédire de manière continue (ex : le niveau d'éveil et de plaisance sur une échelle de 1 à 7) et de manière discrète (ex: une émotion parmi une liste prédéfinie) l'émotion vécue pendant le visionnement de la vidéo (Wang et Ji, 2015). Certaines caractéristiques comme le mouvement, la saillance et le contenu sémantique semblent les plus propices à la compréhension de l'humour (Soleymani et al., 2009).

Le mouvement est souvent utilisé pour évaluer la quantité d'activité dans une vidéo. Le mouvement des objets dans un court clip, tel que mesuré par des vecteurs de mouvement, est positivement corrélé au niveau d'éveil et à une valence négative. Cette relation est généralement expliquée par une association de mouvements rapides avec le danger et l'excitation (Soleymani et al., 2009; Simons et al., 2003; Detenber et al., 1998). La proportion d'éléments saillants dans une vidéo est également utilisée pour trouver les régions d'intérêt et les éléments qui génèrent des émotions fortes dans une vidéo (Zheng et al., 2017; Fan et al., 2017). Un haut niveau de saillance est souvent associé à des émotions plus fortes (valence très positive ou très négative) et une plus grande mémorabilité de l'élément (Fang et al., 2013).

Contrairement à l'humour basé sur le langage (ex: les blagues), la dimension langagière et le contenu sémantique sont beaucoup plus difficiles à étudier lorsque le stimulus humoristique est visuel. Toutefois, plusieurs études commencent à s'y intéresser. Notamment, Chandrasekaran et al. (2016) ont montré que le lien sémantique qui existe entre deux concepts visuels (une grand-mère et un skateboard) est aussi important dans l'humour, si ce n'est pas plus, que les éléments eux-mêmes pris de manière individuelle.

Introduction à l’apprentissage machine

L'apprentissage machine est une technique d’intelligence artificielle où des algorithmes informatiques analysent un ensemble de données afin de déduire des règles qui permettent d'analyser et de prédire de nouvelles situations.

L’algorithme est dit de classification lorsqu’il tente d’attribuer une étiquette discrète à un groupe de données (ex : la vidéo est drôle ou neutre) et de régression lorsque les données prédites sont continues (ex : l’intensité de l’amusement vécu sur une échelle allant de 1 et 100). Les algorithmes d’apprentissage machine peuvent être entraînés avec des données déjà labellisées (entraînement supervisé) afin de trouver un groupe de règles qui différencie les classes ou sans label (entraînement non-supervisé) afin de créer des classes à partir des données.

L’efficacité d’un modèle peut être mesurée de différentes manières. Dans le cas des classifications, on mesure généralement le niveau d’exactitude (*accuracy*) du modèle, c’est-à-dire le taux de valeurs classifié correctement, ou le niveau de précision (*precision*), soit la proportion d'identifications positives qui était effectivement correcte. De plus, il est possible de visualiser l’efficacité d’un modèle de classification grâce aux matrices de confusions. Il s’agit d’une représentation visuelle sous forme de matrice où chaque case représente la proportion de vrai positif, faux positif, vrai négatif et faux négatif. Dans le cadre des régressions, on utilise le plus souvent des techniques telles que l’erreur quadratique moyenne (*mean squared error)* ou la variance expliquée par la régression (e*xplained variance*) pour évaluer l’efficacité du modèle.

Plusieurs algorithmes de classification existent. On retrouve notamment la classification *K-Nearest-Neighbours*, le *Support Vector Machine*, les arbres de décision, et plusieurs autres. Lors du présent mémoire, nous utilisons les Forêts d'arbres décisionnels (*Random Forest*) qui consistent en la création de plusieurs arbres de décision prédisant chacune une valeur. La valeur qui est prédite le plus de fois parmi les arbres de décision est utilisée comme valeur finale prédite par le modèle. Les arbres décisionnels permettent de prendre en compte autant des données continues (ex : niveau d’amusement) que des données discrètes (ex : genre) et permettent d’extraire facilement l’importance des variables dans le modèle final.

Encadré 1. – Introduction à l’Apprentissage Machine

Les caractéristiques de la vidéo sont donc une piste possible pour différencier un contenu humoristique hilarant d’un autre plus ennuyeux. L’utilisation des caractéristiques pour définir l’appréciation de l’humour permettrait de mieux choisir le contenu humoristique à priori d’une étude ou jouer le rôle de variable modératrice dans le cas où l’induction d’amusement via l’humour ne fonctionnerait seulement que chez certains participants.

Prédiction grâce aux expressions faciales

D’un point de vue physiologique, les expressions faciales sont une première opportunité pour évaluer à quel point le participant trouve le contenu drôle. L’avantage des expressions faciales repose sur le fait qu’elles constituent un outil objectif et facilement implémentable dans les expériences. En effet, les études qui utilisent des vidéos humoristiques sont souvent conduites de manière individuelle et à l’aide d’un ordinateur, rendant l’implémentation d’une webcam simple et peu coûteuse. Le sourire est le marqueur prédominant de la majorité des émotions positives, y compris la joie, le soulagement, l'amusement et la gratitude (Hofmann et al., 2017).

Les sourires sont généralement instinctifs et indiquent un sentiment de plaisir. Il implique le muscle zygomatique majeur (AU12), qui attire le coin des lèvres vers les oreilles, et le muscle orbiculaire de l'œil (AU6), qui plisse le coin externe de l'œil en raison de la remontée des joues. Bien que le sourire soit présent dans la majorité des émotions positives, les résultats montrent qu'il est plus intense pendant l'amusement (Hofmann et al., 2017; Herring et al., 2011). Lors d’une méta-analyse explorant la corrélation entre plusieurs émotions et les expressions faciales associées, Dolls et al. (2017) ont constaté que les différentes études rapportent l’amusement comme étant fortement corrélé avec les marqueurs faciaux regroupant le AU6, AU12 et AU25. De plus, contrairement aux autres émotions tel que la joie, l'amusement est la seule émotion avec une forte cohérence entre les études (Dolls et Russell, 2017).

L’utilisation du sourire comporte toutefois certaines difficultés. Les outils actuels nécessitent généralement que les participants soient assis et statiques en face d’une caméra, ce qui peut ne pas convenir à toutes les situations (ex: port d’un masque). De plus, il est possible de sourire ou rire alors que l’état émotionnel de soit pas réellement présent, par exemple, il est possible de forcer un sourire (Krumhuber et Manstead, 2009; Girard et al., 2020). L’utilisation des réactions physiologiques afin de décrire le degré auquel le participant trouve le contenu drôle serait une alternative qui permettrait de pallier ces problèmes.

Prédiction grâce à l’activité cérébrale (EEG)

L’utilisation de l’activité cérébrale est une option intéressante pour prédire à quel point le contenu est perçu comme drôle. Lorsqu'utilisé avec de l'intelligence artificielle, l’activité cérébrale s’est montrée très utile dans la prédiction d’émotions (Liu et al., 2017; Aydın, 2019). Explorer plus précisément la réaction au contenu humoristique grâce au EEG pourrait être bénéfique aux études sur l’humour et l’imagerie cérébrale. De plus, l’appréciation de l’humour induit des changements physiologiques dans l’activité cérébrale qui pourrait être susceptible d’aider à la prédiction de l’intensité.

L’aspect émotionnel de l’humour est principalement associé avec le système dopaminergique de la récompense, dont le cortex préfrontal médian (Iidaka, 2016; Goel et Dolan, 2001; Amir et al., 2013). L’amusement serait également associé à une activation des régions sous-corticales du système de la récompense, incluant le striatum, les noyaux accumbens, l’aire tegmentale ventrale et la substance noire (Bekinschtein et al., 2011; Amir et al., 2013; Campbell et al., 2015; Watson et al., 2007; Mobbs et al., 2003). L’activation du système de la récompense soutient l’idée que l’appréciation d’une situation humoristique est plaisante. Enfin, l’amygdale serait une région importante de l’amusement, ce qui conforte son rôle prédominant dans les processus émotionnels positifs et la détection du niveau de saillance d’un stimulus (Moran et al., 2004; Chan et al., 2012; Bekinschtein et al., 2011; Campbell et al., 2015; Bartolo et al., 2006; Mobbs et al., 2003; Campbell et al., 2015).

Les oscillations cérébrales associées à la perception de l’humour sont rarement étudiées. Barral et al. (2018) ont montré que l’activité gamma était maintenue pendant une plus longue durée après le visionnage d’une vidéo drôle par rapport à un stimulus qui n’était pas considéré drôle. Ainsi, l’activité en gamma pourrait être un prédicteur potentiel du niveau de drôlerie du stimulus. De plus, nous savons que les mouvements moteurs associés au rire sont associés à une diminution de l’activité alpha (8-13Hz) et bêta (14-30Hz) dans les aires motrices (Wang et al., 2017; Mobbs et al., 2003; Caruana et al., 2015).

Projet de Maîtrise

Problématiques & Objectifs

L’humour est un phénomène énormément subjectif et deux participants peuvent réagir complètement différemment à un même contenu humoristique. Malheureusement, ces différences individuelles peuvent affecter la qualité des recherches employant l’humour. En effet, une thérapie par humour pourrait s’avérer non concluante non pas à cause de l’utilisation de l’humour comme thérapie, mais à cause du choix du contenu humoristique qui ne convient pas à tous. À des fins de rigueurs, les études utilisent le plus souvent un stimulus unique, tel que le même extrait de film pour tous ses participants, et ce, au détriment de la variabilité interindividuelle connu dans l’appréciation de l’humour.

Une manière objective de mesurer le niveau d’appréciation humoristique permettrait de prendre en considération la variabilité individuelle des participants sans avoir à sélectionner un contenu personnalisé pour chacun ou modifier grandement la tâche de l’étude. Lors du présent projet de maîtrise, nous avons utilisé des techniques d’apprentissage machine et d’apprentissage profond afin d’explorer trois modalités possibles pour une mesure objective de l'appréciation de l'humour. Nous avons prédit l’appréciation humoristique des participants grâce **(1)** au contenu et aux caractéristiques des vidéos, **(2)** aux expressions faciales (sourire) des participants ainsi que **(3)** à l’aide de l’activité cérébrale tel que mesuré par un EEG commercial.

Articles du mémoire

Le présent mémoire s’inscrit dans le cadre d'une codirection entre les professeurs Karim Jerbi (Département de Psychologie) et Claude Frasson (Département d’informatique et de recherche opérationnelle). Lors de ce projet, nous avons décodé l'intensité de l'amusement de chaque participant à l'aide de techniques d’intelligence artificielle. Une évaluation plus précise et individualisée de l’appréciation de l’humour permettra une sélection de contenu humoristique plus fiable et profitera grandement aux futures implications de l'humour dans les protocoles cliniques et dans la recherche fondamentale. Le mémoire est composé de deux articles, le premier est davantage fondamental, et le suivant est appliqué à l’industrie.

Le premier article (chapitre 2), encadré par le Professeur Jerbi, a permis **(1)** d'étudier de manière fondamentale l'évolution temporelle et incrémentielle de l'appréciation de vidéos humoristiques et de **(2)** prédire l'état émotionnel d'amusement évoqué par les vidéos grâce aux caractéristiques de la vidéo et aux expressions faciales des participants. Dans cette étude, les participants ont regardé et évalué des vidéos humoristiques pendant que leurs expressions faciales étaient enregistrées. Des techniques d'apprentissage machine ont été utilisé afin de prédire l'appréciation humoristique et sa dynamique temporelle.

Le même protocole expérimental a été utilisé en combinaison avec des enregistrements cérébraux EEG dans le développement d'un projet appliqué en entreprise privée. En collaboration avec l'entreprise privée Beam Me Up (BMU) et le Professeur Frasson, nous avons développé un algorithme d'apprentissage profond permettant de prédire en temps réel l'intensité humoristique ressentie grâce à un casque EEG commercial (Emotiv). Le développement et les résultats de l'algorithme sont présentés dans le second article (chapitre 3). Un retour sur les deux études (forces/faiblesses) est détaillé dans la dernière section (chapitre 4) de ce manuscrit.

Chapitre 2 –

Caractérisation des effets comportementaux et physiologiques sur la dynamique temporelle de l'humour (Article 1)

Decoding the intensity of humour appreciation over time.

In preparation for *Physiology & Behavior*

Authors: Gabrielle Toupin[1], Anne-Lise Saive[1], Karim Jerbi[1]

[1] Cognitive and Computational Neuroscience Lab, Psychology Department, University of Montréal, Canada

Corresponding author:

Gabrielle Toupin

E-mail address: gabrielle.toupin@umontreal.ca

Abstract

Humour is a complex cognitive process that can result in a positive emotional state of mirth that has several health benefits. Research and clinical trials use humour as treatment but never actually consider the large variabilities between different individuals. A reliable and objective measure of humour appreciation could benefit future studies employing humour. In this study, 40 participants watched and rated humorous videos while their facial expressions were recorded. For each video clip, we extracted the movement, saliency and two measures of semantic distance. We used physiological reactions and video characteristics at three moments during the trial (beginning, middle, end of the video) to predict how funny (neutral, funny, very funny) the participant would rate the video. Video characteristics were good at predicting between neutral and humorous video but not the intensities of humorous video (accuracy=64.1%). The proportion of smiles was the only feature to fluctuate drastically in time, showing that physiological information is better to identify the humorous moment in the video. Smiles were also better than video characteristics to predict amusement intensities within humorous videos. Furthermore, this study shows that humour appreciation is impacted by how the previous video is perceived. Videos are perceived more funny when the one preceding is funnier ($r=0.477$, $p<0.001$). Smiling associated with a humorous video lasts until the middle of the following one showing that funniness is incremental and can last even when the humorous situation is completed. More than only lasting after the content, the correlation between smiles and funniness rating is stronger after the video ($r=0.703$, $p<0.001$) than during ($r=0.324$, $p<0.001$). This study supports that different humorous content can induce different emotional states. Physiological information such as facial expressions is easy to add to an experiment and would highly benefit further studies using humour.

Keywords: humour, smile, videos characteristics, machine learning, decision tree, time evolution

Introduction

Humour is a complex cognitive process used in social interactions to gently pass criticism, promote social bonding, or even break the ice (Martin, 2007). Finding something funny, *i.e.,* humour elicitation, is often defined as a cognitive and emotional process that takes place in two steps (Suls 1972; Wyer & Collins, 1992). First, humour comprehension, which relies on the detection of incongruous elements in a situation, and the resolution of these incongruities by reinterpreting them in a new and unexpected context. Second, humour elaboration, which represents the emotional response of mirth related to the reinterpretation of these incongruities and includes behavioural markers such as smiling and laughing (Vrticka & al., 2013; Martin, 2007). The emotional response triggered by humour needs to be better characterized because it has been linked to several health benefits. Repeated exposure to humour would allow individuals to better cope with stressful events (Martin, 2004; Martin & Lefcourt, 1983), help prevent the development of cardiovascular diseases and infections (Romundstad et al., 2016), and have long-term improving effects on depression and anxiety (Zhao et al., 2020; Ganz & Jacobs, 2014; Crawford & Caltabiano, 2011; Bennett & al., 2003; Szabo, 2003). However, humour's emotional response remains difficult to consistently induce across people, limiting the robustness of its health benefits and making its use in clinical protocols still very limited (Froehlich & al., 2021; Moran, 1996).

The main challenge of studying humour lies in its variability of appreciation; what makes you roll with laughter may not make your neighbour even smile, and vice versa (Bischetti & al., 2021; Hofmann & al., 2020; Willinger & al., 2017). However, most studies on humour do not assess the level of amusement elicited and simply compare the effects of funny videos to those of neutral or negative videos (Gelkopf & al., 2006; Bennett & al., 2003, Vilaythong & al., 2003). Although comedies are consistently rated as funnier than horror films or romances (Nigbur & Ullsperger, 2020; Wu & al., 2019; Liu & al., 2017), everyone has their own sense of humour that leads to actual differences in intensity of humour appreciation.

In addition, theoretical studies on emotions postulate that the duration of an emotional state highly varies depending on emotion intensity, stimuli duration and the emotional state of participants (for a review, see Verduyn et al., 2015). Furthermore, humour elicitation is a dynamic process eliciting varying levels of amusement over time (Suls 1972, Wyer 1992, Iidaka, 2017), which adds further variability to the assessment of humour appreciation. We believe that by having a better understanding of the temporal dynamics of amusement, it will be easier to predict variability in the appreciation of humour.

This study seeks to provide an appropriate solution for an objective measure of humour appreciation and its temporal dynamic. Instead of looking for a universal humorous cue, we rather propose to decode the intensity of each participant's amusement over time using machine-learning techniques. We believe that a reliable assessment of funniness intensity over time will lead to more reliable humorous content selection, which would highly benefit future implications of humour in clinical protocols and research. In most studies, participants watch humorous clips alone (Boyes & al., 2020; Iidaka, 2017), making it appealing to use facial expressions to predict the reaction to humour. Finding something funny can be characterized by smiling (Soussignan, 2002), which is denoted by the activation throughout the video could also help predict how funny it is perceived when facial expressions are not of two Action Units (AUs) on the Facial Action Coding System: the orbicularis oculi pars orbitalis (AU6) and the zygomaticus major (AU12) (Johnson & al., 2010). Alternatively, what occurs available.

In this paper, we used machine learning techniques to **(1)** predict humour appreciation based on the participant's facial expressions and the video's characteristics and **(2)** explore the temporal dynamics of humour appreciation. Participants watched humorous videos and rated each one on the arousal and valence scale (International affective picture system) to assess their emotional reactions and on a funniness scale. The facial expressions of the participant were recorded throughout the experiment. First, we tried to predict how funny a video was perceived based on the video's characteristics and the participant's facial expression. Good decoding accuracy in either or both features will further support a data-driven approach for funniness evaluation in research and clinical studies. Since funny events usually appear at the end of the video, we expect to have better decoding accuracy during the last section of the video.

Then, we aimed to understand the temporal dynamics of amusement better. We used the proportion of smiles in specific intervals to predict how funny the participant rated the video. Time intervals with higher accuracy should indicate sections of the videos that are more conducive of amusement. We also used the smile during the following video to understand how long the previous video impacted the current one.

Material & Methods

Participants

A total of forty participants (29 women) aged 18-35 years took part in this behavioural experiment. Participants were all in generally good health; three reported a history of neurological or mental illness and were removed from the study. Participants had different levels of education (high school to doctorate). All participants provided written informed consent to procedures approved by the Ethics Committee of the Art and Science Faculty of the University of Montreal (CERAS-2017-18-100-D).

Videos

Stimuli Selection

For this experiment, we created three equal groups of 50 videos, each with three different levels of funniness: neutral, funny and very funny. All videos were cropped to have a length between 8 and 12 seconds, with a global mean of 10 seconds. Black outlines, when present, and sound were systematically removed (videos are available here: https://youtube.com/playlist?list=PLcBTyKtg-JVDx9nAnzD8lnmIvXfH9avnL).

Humorous Stimuli - We used humorous videos from a previous online study to create two groups of humorous videos with different levels of funniness. In this previous online study, 214 participants (141 women) rated 94 humorous video clips, including YouTube video compilations, movie clips and cartoons on arousal, pleasantness and funniness. We used an unsupervised k-means algorithm to devise these humorous videos into two levels of funniness: funny and very funny. For each video, we computed the mean arousal, pleasantness and funniness across participants. To account for the variance in the ratings, we assigned a weight to each video. Videos' ratings with minor variance (i.e. showing a greater agreement between the participants) were assigned more weight in the model.

The first cluster consists of 51 videos with high arousal (m=4.896, STD=0.435), high pleasantness (m=6.125, STD=0.366) and high funniness (m=5.414, STD=0.495). The second cluster contains 43 videos with low arousal (M=4.425, STD=0.410), low pleasantness (M=5.263, STD=0.443) and low funniness (M=4.127, STD=0.445). The two clusters are significantly different in terms of arousal ($t(92) = 5.363$, $p<.001$), pleasantness ($t(92) = 10.330$, $p<.001$) and funniness ($t(92) = 13.132$, $p<.001$).

To obtain the targeted 50 videos of this study, we added and removed some videos. For the very funny groups (51 videos), we dropped the video with the lowest ranking on the three scales. As for the funny group (43 videos), seven new humorous videos validated on four people were added to reach 50 videos.

Neutral Stimuli - An additional control experiment (n=45) was used to collect ratings on 50 neutral videos on arousal and pleasantness (Likert scale from 1 to 9). These videos mostly consist of scenery, people doing everyday activities (ex: walking, biking, meeting) and peaceful animals. All videos were rated with low arousal (mean=4.089; STD=0.733) and neutral pleasantness (nor negative, nor positive) (mean=5.346; STD=0.739). Thus, all neutral videos were used in this study as controls.

Video's Characteristics

For each video included in this experiment, we extracted visual information about the video (Figure 1). We measured the movement and the saliency of each frame since they can help identify the movies' affect, like the genre (ex: action, romance, horror) (Rasheed & al., 2005; Wang & al., 2011). Previous studies in visual humour and machine learning found that semantic content (e.g. tags) can be used to predict the funniness of a scene (Chandrasekaran, 2016; Mahapatra, 2013). With this in mind, we have also created two scores: the semantic distance and the normalized google distance based on keywords automatically generated by a pre-trained algorithm (Video Intelligence) from Google.

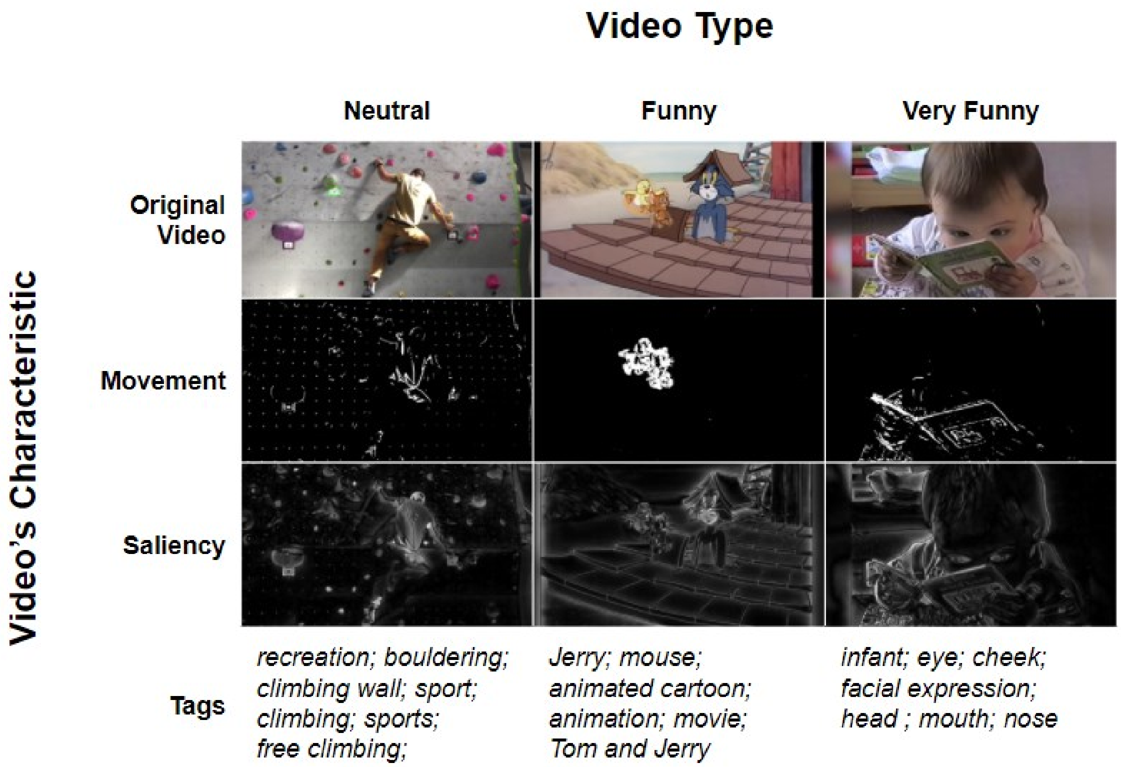

Figure 2. – Visual representation of the video's characteristics extracted

Movement - We extracted the movement for each video with the python library OpenCV (Bradski, 2000). OpenCV is an open-source library specialized in computer vision which allowed us to compute the movement for each frame of video. Since most of our videos have a moving background and were taken mid-action, we opted for a technique where we compare each frame with its previous frame (Wang & Dudek, 2014). We transformed the coloured frame into a gray-scale frame for each video frame, giving us a value between 0 (white) and 100 (black). Then, we computed the absolute value of the difference between the frame *t* and its previous frame *t-1*. For each pixel of the frame, a value above the threshold was considered a moving pixel between frame *t-1* and *t,* while a value under was considered a static object. We selected a strict threshold of 50 to make up for the fact that humorous videos are usually taken in a moving environment, and the filming process is not stable. The final result is a matrix of dimension *[width x height x Frame]* where each data point is a binary value of moving (1) or static (0).

Saliency - The saliency of each frame was obtained using a function implemented by OpenCV (Bradski, 2000). This method calculates the saliency of each video frame based on the spectral residual of an image (Hou & Zhang, 2007). For each video, we get a boolean array *[width x height x Frame]* where each value represents the nature of each pixel: 0 representing a non-salient pixel, and 1 a salient pixel.

Keyword creation - We extracted keywords for each video using an algorithm developed by Google. Video Intelligence (https://cloud.google.com/video-intelligence) is a machine-learning algorithm trained to recognize a large number of objects, places, and actions in videos. The number of keywords extracted for each video varies between 0 and 15, with a mean of 4.66 (STD=3.27) keywords per video. The most popular keywords are *animal*, *pet*, *cat*, *dog* and *sport,* with more than 16 apparitions each. A cleaning process was also required after the keywords acquisition. For each tag obtained, we change the word conjugation to become singular. For example, the word “cats” was transformed to “cat”, which allowed us to remove duplicate words.

***Semantic Distance* -** The semantic distance between two words can be obtained thanks to WordNet, an extensive lexical database of English words. NLTK library (Bird & al., 2009) was used since it is designed to be an interface for WordNet databases. Wu & Palmer’s similarity technique (1994) was used to calculate the relatedness of a pair of keywords. This technique finds the most specific and the closest word between the two (ex: the most specific word between *cat* and *dog* is *domestic animal*) and calculates the number of words between the first and the second words by passing by the most common word. Wu & Palmer (1994) also takes into consideration how specific the word is by calculating how deep the word is in the hypernym tree. This way, for the same distance between two words, generic words such as *animal* and *object* will be considered farther apart than more specific words such as *cat* and *dog*. Since our interest is more specifically in the keywords describing the videos rather than the keywords on their own, we computed descriptive values (mean, standard deviation, minimum value, maximum value and variance) based on all pairs of keywords for a given video.

Normalized Google Distance - Normalized Google Distance (NGD) is a relative semantic distance unit based on how many hits are returned when keywords are searched on google (Cilibrasi &

Vitanyi, 2007). The premise is to use the number of pages where the word “a” and word “b” occur separately and compare with how many pages are returned with both words in it. Thus, if two terms never appear together on the same web page but do occur separately, the NGD between them will be infinite. On the other hand, if both terms always appear together, their NGD score will be zero. For this, we need to use Google’s API to obtain the frequency of each word individually and each pair of words per video. Since Google’s API limits the number of searches, we took up to 10 keywords with the highest confidence level per video. Keywords in common between several videos, such as “cat”, were searched only once to reduce the number of calls. For each video, we computed the NGD score between each pair of keywords. The score of NGD used per video corresponds to the average NGD of each pair of keywords of the video.

Behavioural task and procedure

Participants arrived at the University of Montreal and were required to read, understand, and sign the consent form. Participants were seated comfortably in a small and quiet room with a screen, a mouse, and a keyboard in front of them. At the beginning of the experiment, participants answered demographic questions and completed a short version of the sense of humour questionnaire (SHQ-6; Svebak, 1996). SHQ-6 assesses the habitual relation of participants to humour using three dimensions: the ability to value humour in everyday life situations (*cognitive dimension*), the appreciation of humour in relationships with others (*social dimension*) and the tendency to express emotions in connection with humour (*affective dimension*). Just before starting the experiment, participants evaluated their mood by rating the presence of 20 emotions (10 positives and 10 negatives) on a Likert scale from 1 (very little/not at all) to 5 (extremely) (Positive and Negative Affect Schedule; Watson & al., 1998). At the end of the experiment, the participant answered the same mood questionnaire again.

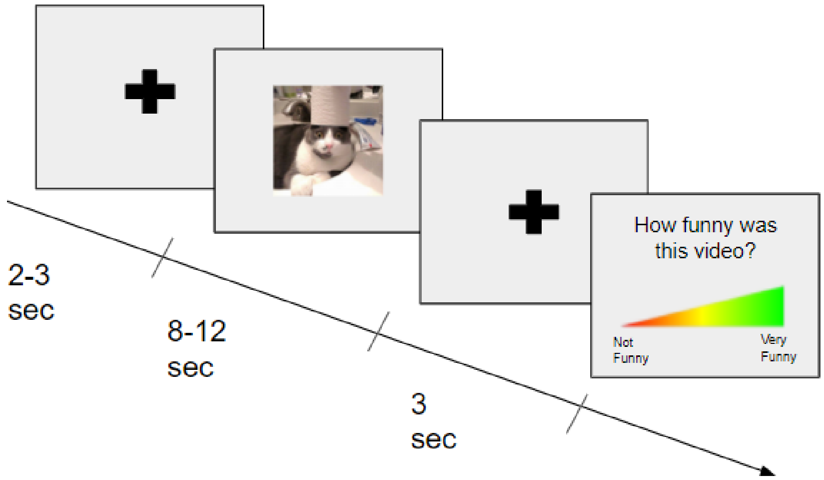

Figure 3. – Schema of a trial. *A trial consists of 2-3 seconds of resting period, followed by a short video of 10 seconds and another resting period of 3 seconds. Participants then rated the video in the following order* ***1)*** *Arousal (calm to excited),* ***2)*** *Pleasantness (negative to positive), and* ***3)*** *Funniness (not funny to very funny)*

The experimental task was created using Psychopy 3 (Peirce & al., 2019) and consisted of three blocks with 50 trials each. Each block was designed with a pseudo-randomized order and included the same amount of neutral, funny videos and very funny videos. We made sure that there were at most three videos of the same type in a row. Each trial consisted of a fixation cross (2-3 seconds), followed by a video (8-12 seconds), another fixation cross (3 seconds), and then video ratings (no time limits) (Figure 2). Participants were asked to rate each video in the following order: "what is your arousal level, i.e. the state of excitement caused by this video in you" (1 = very calm/relaxed and 100 = very excited/stimulated), "what is your level of pleasure, i.e. the emotional state evoked by the video" (1=very unpleasant and 100=very pleasant), and "How funny was the video?" (1 = not funny, and 100 = very funny). All questions were answered using non-graduated scales. During this experiment, the facial expressions of the participant were recorded using a webcam. The experimenter went inside the room between each block and offered the participants to take a couple of minutes to relax.

Physiological Reactions: detection of facial expression with iMotions

Facial expressions were recorded using a webcam during the experiment, and the FACET classifier, included in iMotions software (Stöckli & al., 2018), was used to assess if the participant was smiling or not. FACET automatically estimates facial expressions and reports a score representing the likelihood of 7 emotions and 20 action units (AUs) to be present per frame. In FACET, a positive value indicates that the emotion is likely to be present, while a negative value indicates that the emotion is likely not present. A value of 0 indicates that the emotion has an equal chance of being present or not. For this study, we extracted the proportion of smiles during the video, which is a combination of cheek raise (AU6) and lips corner pulls (AU12). In FACET, those two action units are grouped under the *joy* emotion. For each frame, we assigned a value of 1 (smiling) when the FACET value for *joy* was above 0, and a value of 0 (not smiling) was assigned when the joy was equal or under 0 (Dente & al., 2017). For each time of interest, defined for each analysis below, we computed the proportion of smiles. Thus, a value of 100% means that the person was smiling the entire time, while 0% means that the person was not smiling at all.

Defining three levels of humour appreciation

Further analyses aimed to compare different intensities of funniness while taking personal preferences into account. For each participant, we divided their trials into three equal groups where each group had similar funniness. First, we separated all the videos into: neutral ([0..33e[percentile), funny ([34..66e[percentile) and very funny ([67..100e[percentile). ANOVA and multiple comparisons confirm that the three levels of funniness (neutral, funny and very funny) are all different in terms of funniness ($F(3,50) = 4196$, $p < .001$). Very funny videos (mean=65.88, STD=21.67) are rated as more funny than funny videos (mean=29.23, STD=23.69) which are both more funny than neutral video (mean=5.36, STD=6.63). Additionally, using the same technique, we devised the ratings of each participant into three levels of amusement (low, medium and high) using only humorous videos. We also used the percentile to create three intensities of funniness based for humorous videos only: our low funniness group (mean=25.86, STD= 21.57) is less funny than our moderate funniness group (mean= 50.36, STD= 22.40), which are both less funny than the high funniness group (mean= 71.26, STD= 20.11), ($F(3,33) = 1173$, $p < .001$).

Predicting humour appreciation using machine-learning

In this study, we used a machine-learning approach to predict humour appreciation over time using videos' properties and facial expressions. The present section describes the chosen algorithm, as well as the machine learning pipeline used throughout the study. We ran several classifications (described in sections: *Predicting humour appreciation with video characteristics and proportion of smiles & Temporal Dynamics*) and compared decoding accuracies for different time intervals to characterize the dynamic variations of humour appreciation.

We first try to predict if the participant rated a video as neutral (0-33e percentile of his ratings), funny (33-66e percentile) or very funny (66-100e percentile) based on the characteristic of the video (ex: movement) and the smiles (ex: AU6). We compared each feature's (i.e., *variable*) contribution to the model to see which one helps provide better accuracy. We used this technique on features from the beginning (first third of the video), the middle (second third) and the end (last third) of the video to further define which moment is most decisive in the evaluation of the video. Next, we used the same method on humorous videos only (i.e., without neutral videos), and predictions were made on the three different levels of humour appreciation (low, medium and high funniness). Lastly, we used the same algorithm to predict how the physiological reaction to funniness evolves in time.

Random Forest Classifier

We selected the Random Forest Classifier (RFC) algorithm to use across our study. More specifically, we used the Classification and Regression Trees (CART) version of Scikit-Learn (*https://scikit-learn.org/stable/modules/tree.html#classification-criteria*). RFC combines the results of different Decision Trees and uses the mode to select the final outcome. Decision trees learn the most accurate way to divide the dataset into smaller subsets until it can predict the target value, such as the funniness of our video (*neutral* or *funny*). The condition on which it splits the data (*i.e. node*) is binary (for example: *is the presence of smiles higher than 10%? Yes / No*; *Is the person a woman? Yes / No*). Since a decision is taken at each node, the Random forest classifiers work well with both categorical and continuous values (Strobl & al., 2008).

Furthermore, having explicit conditions for each node allowed us to easily extract the contribution of each feature in the model (Singh & Gupta, 2014). This algorithm's property should be effective in answering which video's characteristic is most beneficial in predicting humour appreciation. A feature (ex: movement) with a higher contribution to the model indicates that it is more critical in decoding the funniness intensity.

Machine Learning Pipeline

How we trained the model is described below, and a schema of the pipeline can be found in Figure 3. First, we randomly split participants into a training group (90% of the participants) and a validation group (10% of the participants). Using the training set, we applied the Scikit-learn random search technique to find the best hyperparameters fitting this data subset. We tried 100 different combinations of hyperparameters (see Table 1 for hyperparameter tuning). For each set of hyperparameters in the random search, we used the leave-one-group-out technique, where each group corresponds to a participant's data. More precisely, we trained a model with the chosen hyperparameters on all the participants minus one and tested it on the unseen participant's data to get the model's accuracy. We repeated this process so that each participant was left out and tested on once. The average of each model represents the accuracy of the final model. From the 100 models tested, we selected the one with the best mean accuracy. We trained the chosen algorithm on the whole training set and tested it on the unseen participants of the validation set excluded at the beginning. When tested on the validation set, we obtained the final accuracy of the model. Since we aimed to predict three categorical data levels, we used permutation tests to determine the chance level for this subset of data. We completed 100 permutations of the model to establish whether our model accuracy predicted above chance level. We then extracted each feature's contribution.

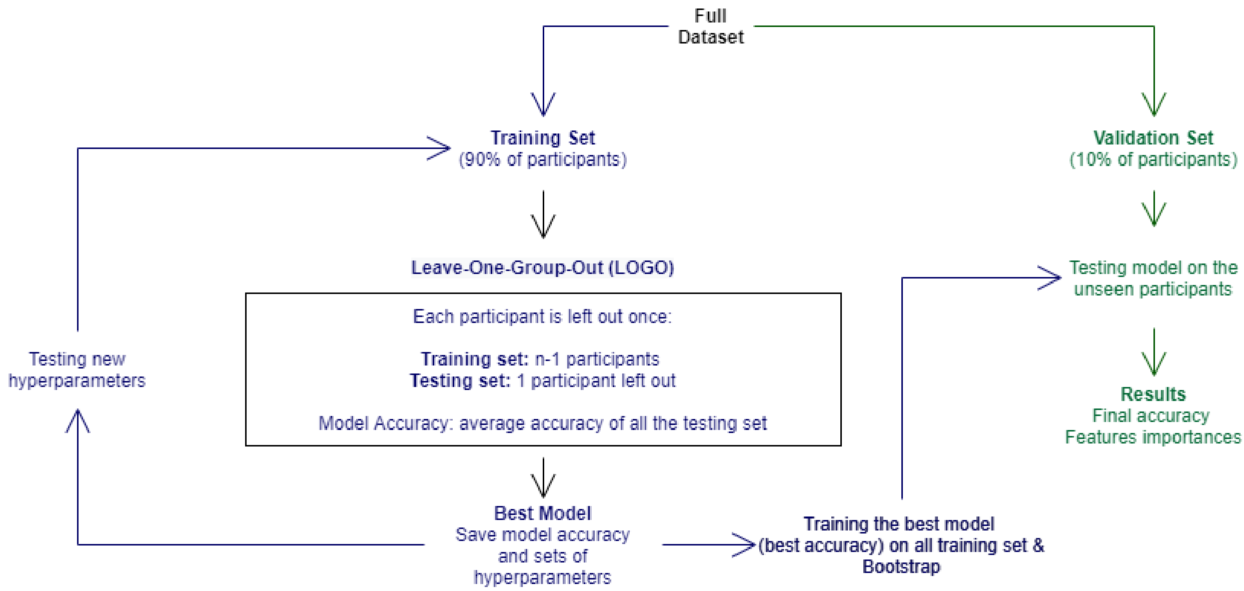

Figure 4. – **Visual representation of the video's characteristics extracted.** *Separation of our dataset into a training set and a validation set; the participant can either be in the training or the validation sets. Leave-One-Group-Out technique was used to tune the hyperparameters of the model. The best model was trained on all the training sets. Scores (e.g. accuracy) are based on the validation set.*

Random Forest Classifier

Hyper-Parameters	Tuning Setting
n_estimators	Value between 100 and 2000 with steps of 10
criterion	['gini', 'entropy']
max_depth	Value between 1 and 100
min_samples_split	Value between 1 and 100
max_leaf_nodes	Value between 1 and 100
max_features	['auto', 'sqrt']
min_samples_leaf	Value between 1 and 200 with steps of 5
bootstrap	[True, False]

Tableau 1. – Hyperparameters tuning of the algorithm

Data Analyses

Behavioural difference in humour appreciation

First, we wanted to compare the behavioural evaluation (*i.e.* ratings of arousal and pleasantness) and physiological reaction (*i.e.* AU6 and AU12) between video categories (*i.e.* neutral, funny and very funny). We used ANOVAs and Tukey's tests to compare each funniness category.

Then, we wanted to observe how our continuous scale of funniness rating correlates with the average presence of smiles during the video and the post-fixation cross. In the first step, this correlation was done using all the videos and in a second step, using only the humorous video. Spearman's correlation was used for all correlations.

Predicting humour appreciation with video characteristics and proportion of smiles

We used the machine learning pipeline described above to define how videos' characteristics explain the intensity of humour appreciation. We divided each video into three intervals based on the video's length: the beginning (first third of the video), the middle (the second third of the video) and the end (last third) of the video. We trained ten models (10x the machine learning pipeline) with different participants for each training and validation set for each time interval. Having multiple models for each time interval guarantees the feature importance obtained is unbiased by the selected participants for the validation set. The features used in this analysis are saliency, movement, semantic distance and normalized google distance for the video's characteristics, as well as the proportion of smiles. Video's properties, like movement and saliency, were specific to the interval of interest, while semantic and normalized google distance were identical for each interval of the same video. This analysis will compare the contribution of video's characteristics and physiological reactions to the model to define which better predict amusement intensity. In the end, we had ten decoding accuracy scores for each time interval, one for each final model created. Additionally, for each feature at each time interval, we have ten contribution values. We use an ANOVA and multiple comparisons to compare the mean decoding accuracy between the three times intervals.

Temporal Dynamics

To explore the temporal dynamics of funniness, we used the proportion of smiles at different points in a trial to predict funniness intensity (neutral, funny, very funny). Since our trials have different lengths, we proportionally divided each trial into smaller intervals of 2-3 seconds each: two sections for the pre-fixation, four sections for the video, two sections for the end-fix and one section for the ratings. Using the machine learning pipeline described above, we computed one model for each interval. The decoding accuracy of each interval was compared to the highest permutation level found across all models (correction for multiple comparisons).

For the time interval to be considered linked to the funniness rating, the accuracy must be higher than the chance level. Furthermore, the higher the accuracy, the more the time interval is key in predicting humour appreciation.

We also examined how long after a video, the participants' smiles were able to decode the type of video they had just seen. This allows us to better understand more precisely the effects of humorous videos over time. We used the proportion of smiles during the following video to predict the funniness of its previous video. If the decoding accuracy is above the chance level, it is considered to be impacted by the funniness of the previous video.

Results

Defining three levels of humour appreciation

Video's ratings

How funny the video was perceived positively impacted the ratings of arousal $F(3,50) = 478$ $p<.001$ and pleasantness $F(3,50) = 961.28$, $p<.001$ (Figure 4). Very funny videos (mean=46.02, STD=23.40) were rated with higher arousal than funny videos (mean=31.59, STD=22.14) which have both higher arousal than neutral videos (mean=24.45, STD=19.55). Very funny videos (mean=67.13, STD=15.72) were rated as more pleasant than funny videos (mean=53.20, STD=15.28) which are also both more pleasant than neutral video (mean=47.64, STD=13.98). In addition, the funniness ratings of the previous video were significantly correlated with the funniness rating of the following one (correlation=0.477, $p<0.001$).

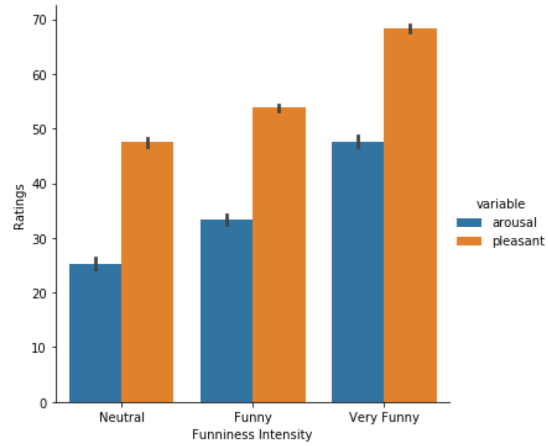

Figure 5. – Evaluation of arousal and pleasantness based on funniness intensity.

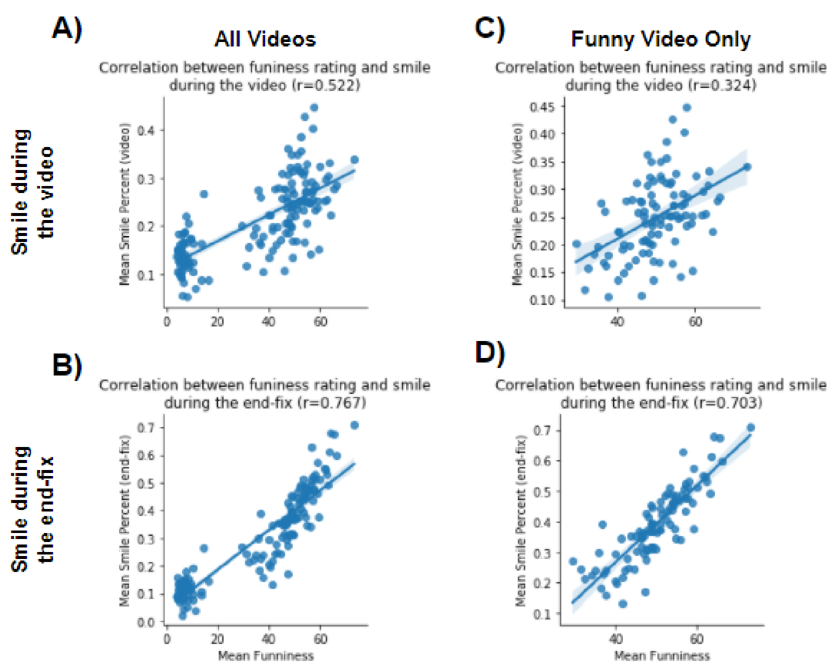

Figure 6. – Relation between funniness and presence of smiles. *Correlation between funniness and the proportion of smiling using neutral and humorous video (A, B) and using only humorous video (C, D). Correlations are shown based on the moment in the video: during the video (A, C) and during the fixation post-video (B, D).*

Smiles

When considering all videos, we observed a high positive correlation (correlation=0.522, p<.001) between the rating of funniness and the proportion of smiles during videos (Figure 5.a). This correlation increased (correlation=0.767, p<.001) when we looked at the proportion of smiles during the resting period following the video (Figure 5.b). Because we noticed that neutral videos were very different from humorous videos, which might have biased correlations, we also conducted the same analysis discarding neutral videos (Fig.5 C, D). We observed similar patterns of correlations, with funniness ratings and smiling significantly correlated both during videos (R = 0.324, p < 0.001), and during the resting period post-video (R = 0.703, p<0.001) (Figure 5.d).

Predicting humour appreciation with video characteristics and proportion of smiles

The results of the funniness prediction based on the video's characteristics and proportion of smiles are presented in this section. We used movement, saliency, semantic distance, normalized google distance and the presence of smiles to predict different levels of funniness. We first considered all videos (a combination of neutral and humorous), followed by humorous videos only.

	Average Feature Contribution (STD)					
Features	**All Videos**			**Humorous video**		
	Beginning	**Middle**	**End**	**Beginning**	**Middle**	**End**
Movement	0.318 (0.008)	0.328 (0.006)	0.220 (0.018)	0.209 (0.033)	0.213 (0.011)	0.171 (0.030)
Saliency	0.291 (0.009)	0.269 (0.007)	0.282 (0.006)	0.248 (0.022)	0.213 (0.019)	0.142 (0.039)
Semantic Distance	0.294 (0.011)	0.248 (0.007)	0.237 (0.006)	0.223 (0.023)	0.191 (0.022)	0.106 (0.036)
NGD	0.059 (0.006)	0.047 (0.002)	0.052 (0.003)	0.089 (0.010)	0.046 (0.007)	0.037 (0.009)
Smile	0.037 (0.021)	0.107 (0.012)	0.209 (0.018)	0.232 (0.053)	0.337 (0.042)	0.543 (0.098)
Mean Accuracy	**62.1% (5.0%)**	**62.3% (5.6%)**	**64.1% (5.4%)**	**41.0% (2.8%)**	**43.9% (3.1%)**	**45.2% (4.6%)**

Tableau 2. – Feature Contribution to models. In this table, we see how each feature contributes to the model at the beginning, the middle and the end of the video. Results are shown for all videos (neutral and humorous) and for humorous videos only. The average value of contribution and the decoding accuracy are shown.

All Types

Model Performance - How well the model was able to predict the funniness was assessed by the accuracy of the model, and permutations were used to confirm that accuracy was higher than the chance level.

The mean accuracy was of 62.1% (std=5.0%) for the beginning of the video, 62.4% (std=5.6%) for the middle part and 64.1% (std=5.4%) for the end. An ANOVA showed no significant difference between the mean accuracy across time ($t(3,10) = 0.417$, $p=.662$). We compared each accuracy score obtained with the chance level as determined by the permutations, and each algorithm was significantly different from the chance level ($p=0.009$).

Feature Contribution - The contribution of each feature was extracted for each of the time clusters. This allowed us to define which features contribute the most to the detection of humour appreciation. Results are described in Table 2 and can be seen in Figure 6.

Smiling is the only feature that has changed drastically over time. At the beginning of the video, smiles contributed very little to the funniness, and it increased for the middle section and again at the end of the video. We looked at the importance of video properties. Semantic anomaly is the only feature that contributed very little to the decoding accuracy at each time in the video, while movement, saliency and semantic distance contributed the most to the decoding accuracy. While looking at the confusion matrix (Figure 7), we saw that our models could accurately predict neutral videos as neutral (accuracy=95%), but most errors appeared between funny (accuracy=50%) and very funny video (accuracy=75%). This would suggest that the neutral videos and the humorous ones differed in terms of video characteristics, but different intensities of humour might not.

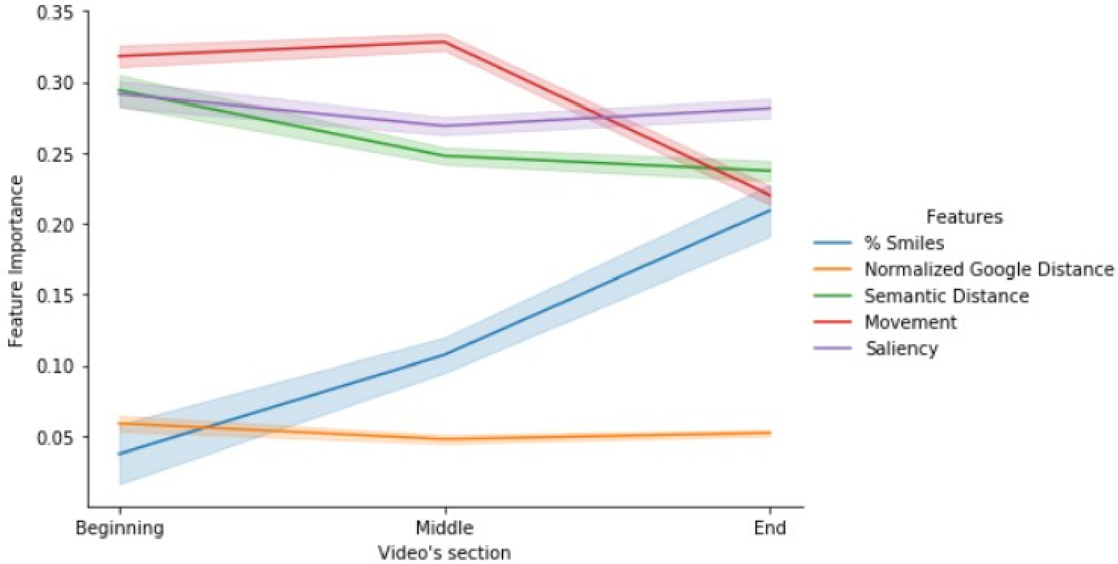

Figure 7. – **Evolution of feature importance when predicting funniness for all videos.** *Evolution in the video (beginning, middle, end) of the contribution of each features to the model decoding accuracy.*

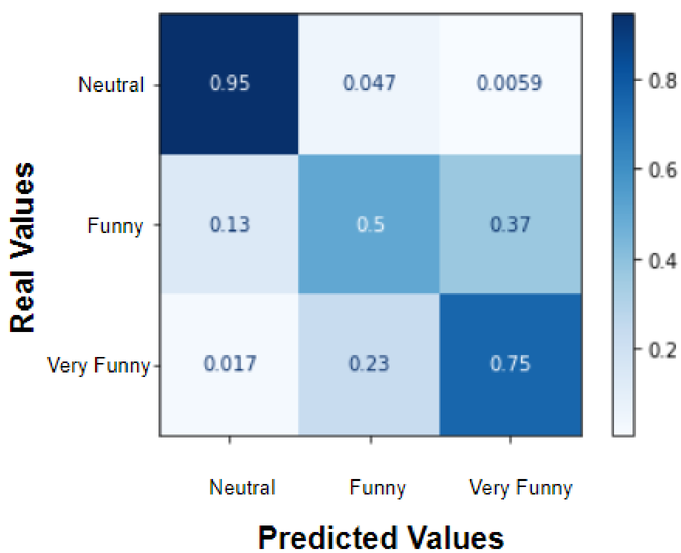

Figure 8. – Confusion Matrix of the best model when selecting all videos. *Confusion matrix from the best model obtained while trying to predict funniness (accuracy=73%). The best model was during the end part of the video.*

Humorous Videos

Models Performance - First, we assessed the quality of our models based on the accuracy. Our models obtained a mean accuracy of 41.1% (STD=2.8%) at the beginning, 44.0% (STD=3.1%) at the middle and 45.2% (STD=4.4%) at the end of the video. All models are above the chance level as defined by permutation tests. An ANOVA showed a significant difference between the three sections of the video ($t\,(3,10) = 3.395$, $p<.001$). Tukey's tests showed that the middle ($p=.05$) and the end ($p=.05$) of the video could predict the funniness intensity better than the beginning of the video.

Feature Contribution - For each feature and time interval, we extracted the feature contribution to the model (Figure 8; Table 2). NGD was lower than the other features for the three intervals. Movement, saliency and semantic distance were stable across time. Furthermore, except for the beginning of the video, where the feature's importance is similar to the video's properties, smiling was more important during the video's middle and end. Confusion matrix (Figure 9) from the best algorithm showed that the algorithm was able to predict the low funniness (accuracy=50%) and high funniness (accuracy=69%) well but not the moderate funniness (accuracy=33%).

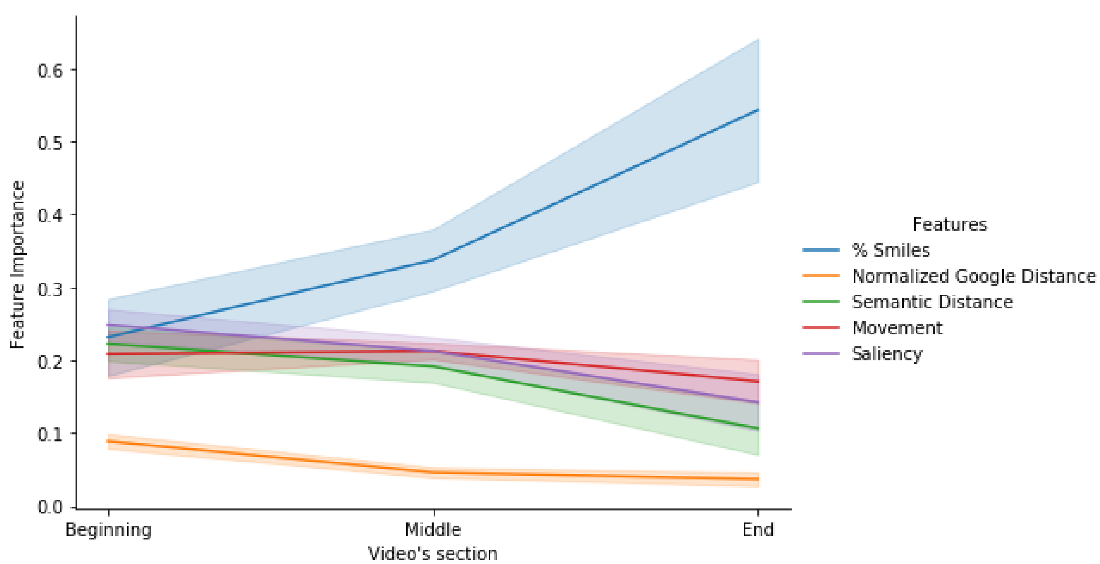

Figure 9. – Evolution of feature importance when predicting the funniness of humorous videos only. *Evolution in the video (beginning, middle, end) of the contribution of each features to the model decoding accuracy.*

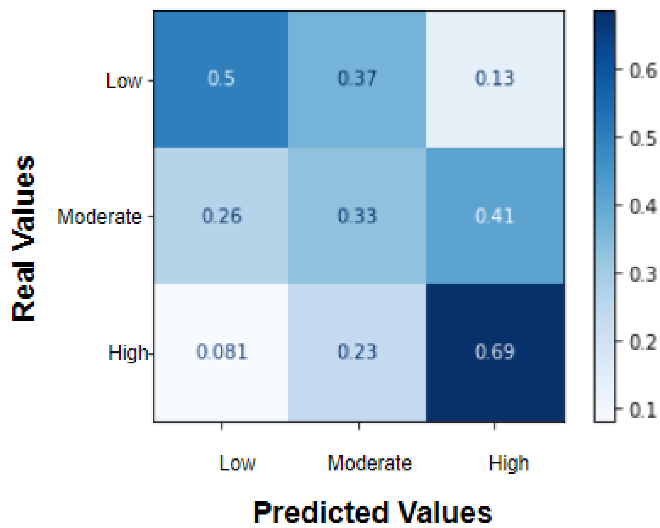

Figure 10. – Confusion Matrix of the best model when selecting humorous video only. *Confusion matrix from the best model obtained while trying to predict funniness (accuracy=51%) on humorous video only. The best model was during the end part of the video.*

Temporal Dynamics

Here, we explored on a more granular level how funniness evolves in time. Using only the physiological response to humour, as denoted by smiles, we tried to predict how funny the trial was perceived at different moments in time. We used the highest permutation score (accuracy=35.9%) as the chance level threshold for all models. An accuracy above this threshold was considered above the chance level, while accuracies under the threshold were considered non-different than chance.

Decoding Accuracy: Time Evolution - Decoding accuracy obtained at each moment is shown in Figure 10. The accuracy level was at chances level during the pre-fixation (prefix-1: accuracy=32.6%, prefix-2: accuracy=32.6%). Starting at the beginning of the video (accuracy=36.3%), the accuracy was above chance level and gradually increased (video-2: accuracy=41.6%, video-3: accuracy=43.9%, video-4: accuracy=49.3%) until it reached its peak at the end of the post-fixation (end-fix-1: accuracy=52.5%, end-fix-2: accuracy=052.5%). The accuracy started to drop when the participant was rating the video (accuracy=45.5%). It decreased slowly (prefix-1: accuracy=45.2%, prefix-2: accuracy=45.2%, video-1: accuracy=41.2%, video-2: accuracy=36.7%) until it reached the chance level in the middle (accuracy=0.353) of the second video and stayed under the chance level for the rest of the trial (video-4: accuracy=31.5%, enfix-1: accuracy=32.5%, endfix-2: accuracy=32.5%, rating: accuracy=33.3%).

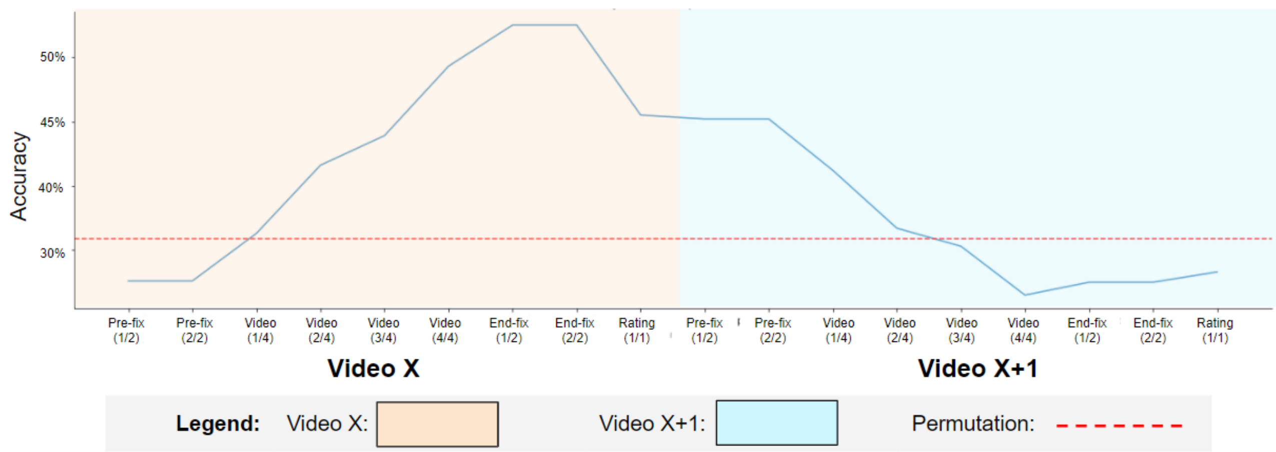

Figure 11. – Evolution of humour appreciation in time. Using Decision Tree Classifier to predict funniness based on how much participant smiled. Evolution of the decoding accuracy in time for the current video (video X) and the following one (x+1). Permutations tests were completed to define the chance level (accuracy=35.9%). Sections division: (section / total number of section)

Discussion

In this paper, we used machine learning techniques to **(1)** predict humour appreciation with video's characteristics and participant's facial expressions and **(2)** explore temporal dynamics of humour appreciation. First, we used the Random Forest Classifier (RFC) at three moments during the video to predict how funny the participants would rate the video. We used the video's characteristics and the proportion of smiles of the participants. Then, we used RFC to predict funniness on smaller time intervals (every 2-3 seconds) to detect more precisely when humour appreciation occurs and how long it lasts.

Behavioural and physiological correlates of humour appreciation

We were able to access the behavioural and physiological reactions to humour appreciation as denoted by the level of funniness perceived by the participants. As expected, neutral and humorous videos manifest differences in terms of behavioural and physiological responses. Similar differences were observed between the two levels of funniness, showing an incremental pattern. The funnier a video was perceived, the more people responded with high arousal and pleasantness. High arousal and pleasantness are associated with positive emotions such as being amused, stimulated, and excited, which correspond to the emotional state of mirth expected from humour (Soleymani & al., 2011; Russel & Lanius, 1984). Lower funniness ratings were closer to the neutral video with lower arousal and neither a positive or negative pleasantness. Since different intensities of humour appreciation were reflected by having different emotional responses, future studies would benefit from including funniness measurements to explain variance in their results.

Furthermore, when a participant rated a video as funnier, he manifested more smiles during and directly after the video. When taking each trial individually, the correlation between the proportion of smiles and funniness was weak, which can be explained by high variance across participants. Some participants did express more smiles than others, with some who have never smiled and some who smiled most of the time (average proportion of smiles per participant: min=1.0%; max=90.1%; mean=25.4%; std=22.9%). Individual differences in manifesting humour appreciation could explain this variance. For example, men (Bischetti & al., 2021; Brody, 2000) and older people (Henry & al., 2013; Thaler & al., 2012; Svebak & al., 2004) tend to be less expressive in their reaction to humour which could lead to a lower proportion of smiles for the same level of funniness.

Best predictive features of humour appreciation

In this second section, we used a data-driven approach to detect funniness in our videos. We explored the possibilities of using either the video’s content and characteristics (movement, saliency, semantic distance & Normalized Google Distance) or the physiological reaction (proportion of smiles) to predict how funny the participant perceived the content. We trained a Random Forest Classifier to predict how the participant will rate the video based on information at three different moments: the beginning, the middle and the end. We tried to predict funniness on all videos (neutral, funny and very funny) and on humorous videos only (low, medium, high funniness). Consistent with our behavioural findings, neutral and humorous videos manifested vast differences in terms of video characteristics and the presence of smiling.

When selecting all the videos, our model could tell above the chance level if the video is neutral or humorous from the moment the video starts (accuracies between 62.1 to 64.1%). The confusion matrix (figure 9) showed that our model accurately differentiates between neutral and all humorous videos but not the intensity of funniness (funny versus very funny). The accuracy stayed similar at every moment in the video, which indicates that the features selected provide the same amount of information at each time interval. This is consistent with the features' contribution, where video properties were most relevant for decoding accuracy and did not fluctuate in time. It would suggest that our neutral videos and our humorous videos were different in terms of video properties but that the two levels of funniness might not be. On the other hand, the physiological reaction of smiling increased in time, which concords with our hypothesis that humour mostly appears at the end of the video. It would suggest that smiling could be more precise to detect the amusement intensity than the video's characteristics while also having a more precise temporal precision.

We then trained the model on humorous videos only. We tried to see if the video's characteristics and proportion of smiles could predict different funniness intensities among the humorous video. First and foremost, the models' accuracy dropped closer to the chance level (accuracy=41%-45%), supporting that our features were mainly applied to classify between neutral and humorous videos. We saw that the smile mainly drives the prediction of funniness (contribution=0.21, 0.39 & 0.53) and that video's properties did not contribute much to the model (contribution=0.08 to 0.27). It confirms that video characteristics did not differ in humorous videos and that smiling would be a better fit to predict funniness intensities among humorous videos. While the accuracy is low, we only tried one machine learning algorithm and further exploration of different algorithms to predict humour appreciation is needed.

Finding a significant difference between our neutral and humorous videos is consistent with numerous studies where movie types are well classified (Hanjalic & Xu, 2005). Emotion intensity, like arousal or discrete emotion, elicited by short clips is still laborious to classify. Soleymani (2009) also had difficulties regarding which multimedia features have the highest correlation with the participant's self-assessment of arousal and pleasantness.

While there might be a greater consensus on what defines a humorous versus a neutral video, different humorous clips might be more specific to individual preferences. Indeed, not being able to classify with high accuracy the funniness intensity raises the idea that different levels of movement, saliency and semantic could be favoured by different people. Both individual characteristics such as demographic (e.g. gender, age) and psychological scale (e.g. Martin (2003) humour style questionnaire) should be included in future research using videos characteristics to predict amusement.

Temporal dynamics of humour appreciation

Unlike the rating of funniness at the end of the video, which offers a global understanding of amusement during the video, the proportion of smiles allows to capture the amusement at every point during the video. Thus, we used the proportion of smiles to predict the funniness intensity at specific moments in time to explore on a granular level how funniness evolves during the video.

Before the video started, the participant did not know if he would see a humorous or neutral video. Thus, we expected that it would be impossible to determine above the chance level how he would evaluate the video's funniness. Surprisingly, from the moment the video starts, the proportion of smiles predicted above the chance level the funniness of the video, suggesting that participants started to find funny elements in the videos even at the beginning. The funniness of the video increased the more we advanced in the video, showing that the video's funny element tends to be nearer the end of the video. Those results are in line with humour theories, where the first part is the settings, followed by the detection and resolution of incongruous detection. Once we resolve the cognitive phase, emotional reaction and appreciation emerge (Martin, 2007).

Physiological reactions to funniness, as denoted by smiles, were particularly more present during the end of the video and directly after the video (post-fixation & rating). This is important for physiological studies using humour since it supports that the period after the humoristic stimuli is as essential, if not more, than the period during the presentation of the stimuli. It further supports studies, such as Barral & al. (2017), for selecting a special window composed of the last 2 seconds of the cartoon and the next 5 seconds to identify brain activity linked to humour.

We also used the smile during the following video to understand how long the previous video impacted the one currently seen. Using the smiles during the following video, we tried to predict the previous video's type. Intervals with accuracy above the chance level indicate that the interval is still affected by the previous video. This technique allowed us to observe how long humour appreciation lasts after viewing the video. While the fixation before the video cannot predict how funny the video will be, it could predict how the previous video was rated.

Conclusion

Humour appreciation, as defined by how funny we find something, varies tremendously between different contents and individuals. In this study, we aimed to find an objective method to detect the intensity of humour appreciation. We used machine learning techniques to predict how each participant will rate a humorous video based on the video's characteristics (movement, saliency and semantics) and the participant's smile. The video's content and characteristic differed significantly between humorous and neutral videos, making it a great choice to validate if the video is humorous or not when compared to neutral. On the other hand, video characteristics did not differentiate intensities of funniness in humorous videos. This could be explained by individual preferences for different characteristics in the video (ex: preference for high movement), which were not considered in the machine learning pipeline used in this study. Physiological reaction, such as the proportion of smiling, was better than the video's characteristic to predict the funniness intensities of humorous video.

Furthermore, smiling is the only feature used in this study that fluctuates during the video, showing greater granularity in time to detect humour appreciation. This higher granularity allowed us to explore how humour appreciation evolves during and across trials. Our results showed that humour appreciation is present during the entire video and is stronger nearer the end of the video. Smiles were also at their peak directly after viewing the video and slowly decreased until the middle of the subsequent trial. This was concordant with our behavioural results, where participants rated a video as funnier when the preceding one was also funnier.

Strengths & Limitations – In this study, we used state-of-the-art techniques to predict how participants reacted to humorous videos. We used a large subset of videos pre-validated to ensure that every participant found something funny. Moreover, we were able to characterize the temporal dynamics of humour appreciation.

On the other hand, we believe that higher decoding accuracy could be reached if demographics and individual information, such as the sense of humour of the participant, could be included in the ML pipeline. It could help define preferences in terms of favourite video characteristics and define an average proportion of smiles for different types of participants. Furthermore, while using Google API *Video Intelligence* to generate keywords for our video is easy to implement in future studies, a better understanding of how semantics impact visual humour is needed and would benefit from having keywords chosen by participants. In this study, we selected only two methods to predict humour appreciation: video characteristics and participants' facial expressions. However, other physiological measures such as brain activity, skin conductance, heartbeat and brain activity could also be potential alternatives to detect humour appreciation.

Impacts on future studies – The ability to objectively assess humour appreciation in researches and clinical trials using humour is key. In this study, using facial expression was the best feature to decode the intensity of humour appreciation. This data-driven approach to measure how the participant finds the content funny would be an excellent add-on to include in clinical studies and fundamental research. It would allow the use of humour appreciation as a moderator variable. Furthermore, the fundamental understanding of how humour appreciation evolves in time will allow a better content selection and better selection of time interest in futures physiological studies.

References

Barral, O., Kosunen, I., & Jacucci, G. (2017). No need to laugh out loud: Predicting humor appraisal of comic strips based on physiological signals in a realistic environment. ACM Transactions on Computer-Human Interaction (TOCHI), 24(6), 1-29.

Bennett, M. P., Zeller, J. M., Rosenberg, L., & McCann, J. (2003). The effect of mirthful laughter on stress and natural killer cell activity.

Bird, S., Klein, E., & Loper, E. (2009). Natural language processing with Python: analyzing text with the natural language toolkit. “O’Reilly Media, Inc.”.

Bischetti, L., Canal, P., & Bambini, V. (2021). Funny but aversive: a large-scale survey of the emotional response to COVID-19 humor in the Italian population during the lockdown. Lingua, 249, 102963.

Boyes, M. E., Clarke, P. J., & Hasking, P. A. (2020). Relationships between dispositional and experimentally elicited emotional reactivity, intensity, and perseveration. Personality and Individual Differences, 152, 109573.

Bradski, G. (2000). The opencv library. Dr Dobb’s J. Software Tools, 25, 120-125.

Brody, L. R. (2000). The socialization of gender differences in emotional expression: Display rules, infant temperament, and differentiation. Gender and emotion: Social psychological perspectives, 2.

Chandrasekaran, A., Vijayakumar, A. K., Antol, S., Bansal, M., Batra, D., Zitnick, C. L., & Parikh, D. (2016). We are humor beings: Understanding and predicting visual humor. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 4603-4612).

Cilibrasi, R. L., & Vitanyi, P. M. (2007). The google similarity distance. IEEE Transactions on knowledge and data engineering, 19(3), 370-383.

Crawford, S. A., & Caltabiano, N. J. (2011). Promoting emotional well-being through the use of humour. The Journal of Positive Psychology, 6(3), 237-252.

Dente, P., Küster, D., Skora, L., & Krumhuber, E. (2017, April). Measures and metrics for automatic emotion classification via FACET. In Proceedings of the Conference on the Study of Artificial Intelligence and Simulation of Behaviour (AISB) (pp. 160-163).

Fang, Y., Lin, W., Chen, Z., Tsai, C. M., & Lin, C. W. (2013). A video saliency detection model in compressed domain. IEEE transactions on circuits and systems for video technology, 24(1), 27-38.

Froehlich, E., Madipakkam, A. R., Craffonara, B., Bolte, C., Muth, A. K., & Park, S. Q. (2021). A short humorous intervention protects against subsequent psychological stress and attenuates cortisol levels without affecting attention. Scientific Reports, 11(1), 1-10.

Ganz, F. D., & Jacobs, J. M. (2014). The effect of humor on elder mental and physical health. Geriatric Nursing, 35(3), 205-211.

Gelkopf, M., Gonen, B., Kurs, R., Melamed, Y., & Bleich, A. (2006). The effect of humorous movies on inpatients with chronic schizophrenia. The Journal of nervous and mental disease, 194(11), 880-883.

Hanjalic, A., & Xu, L. Q. (2005). Affective video content representation and modeling. IEEE transactions on multimedia, 7(1), 143-154.

Henry, J. D., Phillips, L. H., Ruffman, T., & Bailey, P. E. (2013). A meta-analytic review of age differences in theory of mind. Psychology and Aging, 28(3), 826.

Hofmann, J., Platt, T., Lau, C., & Torres-Marín, J. (2020). Gender differences in humor-related traits, humor appreciation, production, comprehension,(neural) responses, use, and correlates: A systematic review. Current Psychology, 1-14.

Hou, X., & Zhang, L. (2007, June). Saliency detection: A spectral residual approach. In 2007 IEEE Conference on computer vision and pattern recognition (pp. 1-8). Ieee.

Iidaka, T. (2017). Humor appreciation involves parametric and synchronized activity in the medial prefrontal cortex and hippocampus. Cerebral Cortex, 27(12), 5579-5591.

Johnson, K. J., Waugh, C. E., & Fredrickson, B. L. (2010). Smile to see the forest: Facially expressed positive emotions broaden cognition. Cognition and emotion, 24(2), 299-321.

Liu, Y. J., Yu, M., Zhao, G., Song, J., Ge, Y., & Shi, Y. (2017). Real-time movie-induced discrete emotion recognition from EEG signals. IEEE Transactions on Affective Computing, 9(4), 550-562.

Mahapatra, A. (2013). A computational approach to detection of conceptual incongruity in text and its applications.

Martin, R.A. (2007). The Social Psychology of Humor. In The Psychology of Humor, (Elsevier), pp. 113–152.

Moran, C. C. (1996). Short-term mood change, perceived funniness, and the effect of humor stimuli. Behavioral medicine, 22(1), 32-38.

Nigbur, R., & Ullsperger, M. (2020). Funny kittens: Positive mood induced via short video-clips affects error processing but not conflict control. International Journal of Psychophysiology, 147, 147-155.

Peirce, J. W., Gray, J. R., Simpson, S., MacAskill, M. R., Höchenberger, R., Sogo, H., Kastman, E., Lindeløv, J. (2019). PsychoPy2: experiments in behavior made easy. Behavior Research Methods. 10.3758/s13428-018-01193-y

Rasheed, Z., Sheikh, Y., & Shah, M. (2005). On the use of computable features for film classification. IEEE Transactions on Circuits and Systems for Video Technology, 15(1), 52-64.

Romundstad, S., Svebak, S., Holen, A., & Holmen, J. (2016). A 15-year follow-up study of sense of humor and causes of mortality: The Nord-Trøndelag Health Study. Psychosomatic Medicine, 78(3), 345-353.

Russell, J. A., & Lanius, U. F. (1984). Adaptation level and the affective appraisal of environments. Journal of Environmental Psychology, 4(2), 119-135.

Singh, S., & Gupta, P. (2014). Comparative study ID3, cart and C4. 5 decision tree algorithm: a survey. International Journal of Advanced Information Science and Technology (IJAIST), 27(27), 97-103.

Soleymani, M., Chanel, G., Kierkels, J. J., & Pun, T. (2009). Affective characterization of movie scenes based on content analysis and physiological changes. International Journal of Semantic Computing, 3(02), 235-254.

Soleymani, M., Lichtenauer, J., Pun, T., & Pantic, M. (2011). A multimodal database for affect recognition and implicit tagging. IEEE transactions on affective computing, 3(1), 42-55.

Soussignan, R. (2002). Duchenne smile, emotional experience, and autonomic reactivity: a test of the facial feedback hypothesis. Emotion, 2(1), 52.

Stöckli, S.; Schulte-Mecklenbeck, M.; Borer, S. & Samson, A.C. (2018) Facial expression analysis with AFFDEX and FACET: A validation study. Behavior Research Methods, 50 (4), 1446-1460.

Strobl, C., Boulesteix, A. L., Kneib, T., Augustin, T., & Zeileis, A. (2008). Conditional variable importance for random forests. BMC bioinformatics, 9(1), 1-11.

Suls, J. M. (1972). A two-stage model for the appreciation of jokes and cartoons: An information-processing analysis. The psychology of humor: Theoretical perspectives and empirical issues, 1, 81-100.

Svebak, S. (1996). The development of the Sense of Humor Questionnaire: From SHQ to SHQ-6.

Svebak, S., Martin, R. A., & Holmen, J. (2004). The prevalence of sense of humor in a large, unselected county population in Norway: Relations with age, sex, and some health indicators. Humor, 17(1-2), 121-134.

Szabo, A. (2003). The acute effects of humor and exercise on mood and anxiety. Journal of leisure research, 35(2), 152-162.

Thaler, A., Posen, J., Giladi, N., Manor, Y., Mayanz, C., Mirelman, A., & Gurevich, T. (2012). Appreciation of humor is decreased among patients with Parkinson’s disease. Parkinsonism & related disorders, 18(2), 144-148.

Verduyn, P., Delaveau, P., Rotgé, J. Y., Fossati, P., & Van Mechelen, I. (2015). Determinants of emotion duration and underlying psychological and neural mechanisms. Emotion Review, 7(4), 330-335.

Vilaythong, A. P., Arnau, R. C., Rosen, D. H., & Mascaro, N. (2003). Humor and hope: Can humor increase hope?. Humor, 16(1), 79-89.

Vrticka, P., Black, J. M., & Reiss, A. L. (2013). The neural basis of humour processing. Nature Reviews Neuroscience, 14(12), 860.

Wang, B., & Dudek, P. (2014). A fast self-tuning background subtraction algorithm. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops (pp. 395-398).

Wang, J., Li, B., Hu, W., & Wu, O. (2011, May). Horror video scene recognition via multiple-instance learning. In 2011 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) (pp. 1325-1328). IEEE.

Watson, D., Clark, L. A., & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: The PANAS scales. Journal of Personality and Social Psychology, 54, 1063-1070.

Willinger, U., Hergovich, A., Schmoeger, M., Deckert, M., Stoettner, S., Bunda, I., ... & Auff, E. (2017). Cognitive and emotional demands of black humour processing: the role of intelligence, aggressiveness and mood. Cognitive processing, 18(2), 159-167.

Wu, Y., Gu, R., Yang, Q., & Luo, Y. (2019). How do amusement, anger and fear influence heart rate and heart rate variability?. Frontiers in Neuroscience, 13, 1131.

Wu, Z., & Palmer, M. (1994). Verb semantics and lexical selection. arXiv preprint cmp-lg/9406033.

Wyer, R. S., & Collins, J. E. (1992). A theory of humor elicitation. Psychological review, 99(4), 663.

Zhao, J., Yin, H., Wang, X., Zhang, G., Jia, Y., Shang, B., ... & Chen, L. (2020). Effect of humour intervention programme on depression, anxiety, subjective well-being, cognitive function and sleep quality in Chinese nursing home residents. Journal of Advanced Nursing, 76(10), 2709-2718.

Chapitre 3 –

Prédiction de l'intensité de l’amusement basée sur l'activité cérébrale (Article 2)

Prediction of amusement intensity based on brain activity

Published in *Frontiers in Artificial Intelligence and Applications (FAIA)*

DOI: 10.3233/FAIA210099

Authors: Gabrielle Toupin[1], Mohamed Sahbi Benlamine[2], Claude Frasson[2]

[1] Cognitive and Computational Neuroscience Lab, Psychology Department, University of Montréal, Canada

[2] Department of Computer Science and Operations Research, University of Montréal, Canada

Corresponding author:

Gabrielle Toupin

E-mail address: gabrielle.toupin@umontreal.ca

Abstract

Amusement can help modulate psychological disorders and cognitive functions. Unfortunately, algorithms classifying emotions still combine multiple positive emotions into a unique emotion, namely joy, making it hard to use amusement in a real-life setting. Here we train a Long-Short-Term-Memory (LSTM) on electroencephalography (EEG) to predict amusement on a categorical scale. Participants (n=10) watched and rated 120 videos with various funniness levels while their brain activity was recorded with an Emotiv Headset. Participants' ratings were divided into four bins of amusement (low, medium, high & very high) based on the participant's ranking's percentile. Nested cross-validation was used to validate the models. We first left out one video from each participant for the final model's validation, and a leave-one-group-out technique was used to test the model on an unseen participant during the training phase. The nested cross-validation was tested on sixteen different videos. We create an LSTM model with a hidden layer of 100 neurons, with a batch size of 256 and an input layer of 14x128 (electrodes x 1 sec of recording) and four nodes representing the different levels of amusement. The best model obtained during the training phase was tested on the unseen video. While the level of accuracy between the validation videos varies slightly (mean=57.3%, std=13.7%), our best model obtained an accuracy of 82,4%. This high accuracy supports the use of brain activity to predict amusement. Moreover, the validation process we design conveys that models using this technique are transferable across participants and video.

Keywords: amusement, LSTM, EEG, emotiv, emotions, humour

Introduction

Context & Motivation

Humour is a social behaviour that allows people to break the ice, relax the atmosphere, or gently pass a criticism (Martin, 2007). It is a complex cognitive process that can result in an emotional state of amusement and can trigger laughter (Vrticka & al., 2013). Research in positive psychology induces amusement to modulate psychological disorders, such as schizophrenia and depression (Gelkopf & al., 1993; Hirosaki & al, 2013; Zhao & al., 2020). This positive emotion can also benefit cognitive functions such as memory (Zhao & al., 2020; Savage & al., 2017). In addition to having different research uses, amusement differs from joy in terms of facial expressions, physiological signals, and feelings (Herring & al., 2011; Liu &al., 2017). Nevertheless, predicting emotions still widely combines these positive emotions together (Kim & al., 2019). Only a handful of studies can predict different positive emotions (Liu & al., 2017; Aydın, 2019). Thus, a better understanding and prediction of amusement would benefit research using amusement as a regulator, clinical research, and new technologies.

The development of new algorithms to predict emotions in artificial intelligence is on the rise. Those algorithms are trained to predict emotions based on facial expressions, electroencephalography (EEG) or physiological signals, such as electrocardiography. Algorithms using facial expressions to predict emotions can be complicated when used in real-world applications and experiments. First, using a filming process requires specific settings where the participant always faces a camera, making it notably difficult for moving subjects and situations where faces are hidden (e.g., virtual reality headset, mask). So far, algorithms based on artificial intelligence, like Emotient, are better than humans at classifying basic emotions when they are typical, exaggerated and static. However, the accuracy drastically drops when used on spontaneous, dynamic and mixed emotions (Stöckli & al., 2018; Duan & al., 2013; Craik & al., 2019). While there is more work to be done in this area, using brain activity and physiology might be a better choice to train algorithms to predict emotion since it does not require the participant to be static in front of the camera, and physiological signals cannot be intentionally controlled. With the use of new technologies like Emotiv (https://www.emotiv.com/), where the headset is affordable, requires minimum setup and connects via Bluetooth, new setup experiments and real-life applications are conceivable.

Researchers use different estimators and features to train algorithms to detect emotion based on EEG signals. If we look at the machine learning side, studies use estimators such as support vector machines, Naive Bayes and K-nearest neighbours to classify emotions (Liu & al., 2017, Aydın, 2019; Duan & al., 2013). When looking at deep learning, there is no consensus on which algorithm is best for emotion classification (Craik & al., 2019). In their study, Alhagry (2017) reached an accuracy score over 85% with a Long-Short Term Memory algorithm (LSTM) to predict the intensity of arousal and valence of the emotion based on raw EEG. Using LSTM algorithms is promising since it can learn from complex data and predict both on a continuous and categorical scale. Feeding raw EEG data allows us to create algorithms that do not require transformed data, which takes time to compute.

Furthermore, LSTM can take more information into account than classical machine learning techniques, meaning that even some artifacts or movements detected by the EEG headset could help define the amusement intensity. Therefore, we propose to train an LSTM algorithm to predict amusement intensity with EEG data acquired with the Emotiv headset. This study brings new insights into the prediction of emotion intensity and amusement.

Objectives

As the first and primary objective, we want to confirm that EEG data collected with an Emotiv Epoch can predict the amusement level of the participant. We will train an LSTM to predict the categorical score of amusement (low, medium, high, very high) based on one second of brain activity and 14 electrodes. Our second objective is to design an innovative pipeline that assures our model's transferability to new participants and new visual content.

Methods

Participants

Ten participants (7 women, 3 men) were recruited for this experiment. They were approached on social media and were offered monetary compensation in exchange for their participation.

Recruited participants were between 18-30 years old (mean=24, STD=3.88) and had similar education, standard or corrected-to-normal vision and no neurological or psychological disorders. One participant had to be excluded since he never completed the active task.

Material

One hundred twenty video clips were used as stimuli. These video clips were selected in two steps. First, undergrad volunteers selected short portions of humorous and neutral videos from movies, short clips and video compilations. A total of 50 neutral videos and 100 humorous videos were selected. The videos were cropped to have a length between 8 to 12 seconds (mean of 10 seconds). Furthermore, black outlines and the sound, when they were present, were removed. Second, in a previous study, forty participants watched and rated every video on the following scales: arousal, valence and funniness. To reduce the amount of times participants would have to wear the headset: we selected 120 videos based on the following method. First, we performed a K-Means clustering with three clusters on the behavioural ratings to define three intensities of amusement: neutral (low on all scales), funny (average on all scales) and very funny (high on all scales). Creating three intensities of amusement will ensure that videos are diverse in terms of amusement and will help define a pseudo-randomized to show the videos. Then, for the three intensities of funniness, we selected the 40 videos with the closest distance to the centroid to represent the final video used in this study.

Procedure

Participants arrived at the Functional Neuroimaging Unit and were required to read, understand, and sign the consent form. Participants were seated comfortably in a Faraday Cage. There was a screen, a mouse, and a keyboard in front of them to perform the experiment. The task used in this study was created with Psychopy 3 (Peirce & al., 2019) and consisted of four blocks with 30 trials each. Each bloc was designed with a pseudo-randomized order and included ten neutral videos, ten funny videos and ten very funny videos. We made sure that there were at most three videos of the same type in a row. A single trial consists of a fixation cross (2-3 seconds), followed by a video (8-12 seconds), another fixation cross (3 seconds) and a single question ("how funny was this video") (see Figure 1). The question was on a scale of 1 (not funny) to 100 (very funny) and was answered with the mouse. An active task was used while watching the video to keep the participant engaged in the task. The participant was asked to press the spacebar on the keyboard when he believed the video was supposed to be humorous.

The first part of the experiment consisted of a practice block where the participant got familiar with the trial design in the experimenter's presence. The participant was free to ask any question about the trial, and the experimenter made sure that the task was understood. Prior to the beginning of the test, participants also answered a short questionnaire about how they are feeling right now based on 20 emotional statements (Positive and Negative Affect Schedule; Watson & al., 1998). Then, a resting state (6 minutes) was measured. During the resting state, the participant was asked to look at the cross in the screen's center and stay neutral while we recorded its brain activity when no explicit task is not being performed. The participant was then ready to start the four real blocks. Once the participant was ready, he could press the keyboard's space button to start the block. The experimenter went inside the room between each block and ensured the participant was still in good shape to proceed with the task. It was recommended to take a couple of minutes to relax between each block. After all the videos were viewed, the participant completed another resting state and a final emotional questionnaire.

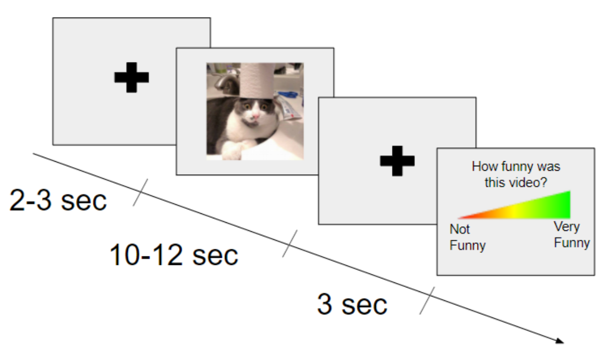

Figure 1. – Behavioural Task: Trial

EEG recording

The Emotiv Epoc headset was used to collect electrical activity during the task. Brain activity was recorded from 14 electrodes (AF3, F7, F3, FC5, T7, P7, O1, O2, P8, T8, FC6, F4, F8, and AF4) (Figure 2) with two reference nodes located behind the ears. The generated data are in microvolt (μV) with a sampling frequency of 128Hz. Electrodes were moisturized with a saline solution to maintain electrode impedance under the software's required level. Impedance was checked during the initial installation, followed by a rechecked before the start of each block.

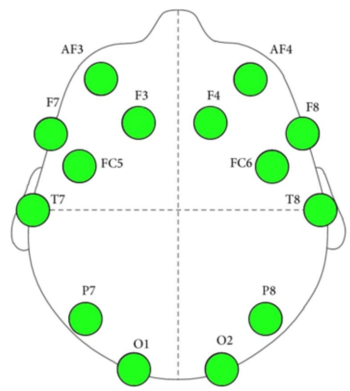

Figure 2. – Disposition of electrodes in the Emotiv Epoch

Data Preparation

Participant Evaluation of Amusement

Only humorous videos were used to develop the model. Participants rated humorous video on highly different scales: some participants have their lowest rating around 0 while others have nothing below 58. Similarly, highest values rated by participant varies between 61 and 100. Since the rating interval differs between the participants, we scaled the rating between 0 and 1 for each participant. The participant's lowest value was converted to 1, his highest value to 100 and every value in between was scaled proportionately. We computed the user's amusement rating by dividing the rating scale into four intervals: [0..0.25[for low amusement, [0.25..0.50[for medium amusement, [0.50..0.75[for high amusement and [0.75..1[for very high amusement which was used to label the dataset.

Time of Interest

Funniness appears mostly at the videos' end (Toupin & al., 2021; Barral, & al., 2017), leading us to choose the video's end as the time of interest for humorous videos (Figure 3). More precisely, if the participant pressed the button to indicate that it is indeed a humorous clip, we only used EEG data between the button press and the end of the video and assigned the user's reported amusement rating. On the other hand, if no button was pressed, mainly seen in less funny videos, we assumed that the reported funniness was stable across the video and used EEG data associated with the full video's length.

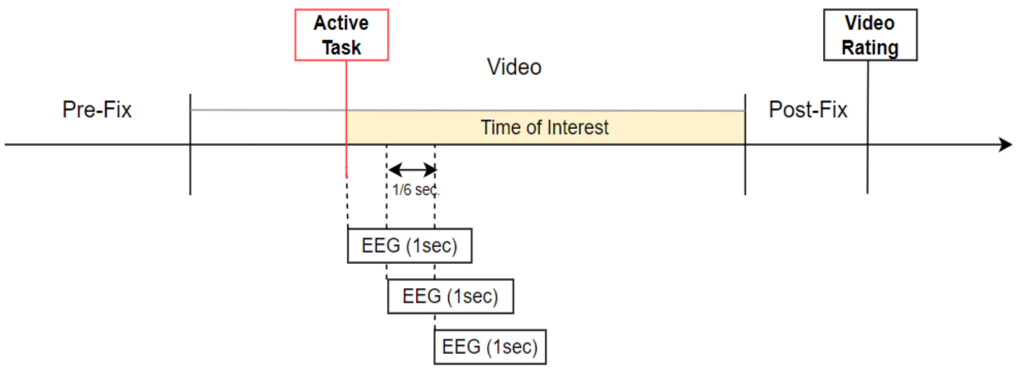

Figure 3. – Time of Interest during the trial

EEG Data Cleaning

EEG data collected were cleaned via Emotiv Software and Python's MNE library. The first cleaning part is done automatically by the Emotiv Software, where it uses a 5th-order digital Sinc to filter between 0.2 Hz and 45 Hz. Plus, it uses a Notch filter at 60Hz since it is the frequency band for North America's electricity. Emotiv software also automatically removes most of the eye's blinks and heartbeat from the signal.

Additionally, we complemented the cleaning from Emotiv Software with an additional process done with Python's library MNE. To validate that all eyes and cardiac artifacts were well removed, we decomposed the EEG signal using an Independent Component Analysis (ICA). ICAs that were strongly correlated with either eye blinks or heartbeat were removed from the signal. ICAs did not remove any artifacts from the signals, showing that they were already removed with the

Emotiv software. Finally, we manually observed the signal of each participant and annotated the noisy parts of the signal. Epochs with those annotated parts were not used in further analysis.

Features and Labels

Our model will attempt to predict the user’s amusement rating from 1 second of EEG data from all electrodes. We used a data matrix of shape 128x14, which holds 1 second of recording for each of the 14 electrodes. This one second of recording was associated with the rating of the participant for this specific video (low, medium, high, very high amusement). For the length of the trial's time of interest, we move the data matrix 1/6 second in time and assign the participant's rating to the data matrix.

Model Training and validation

Deep-Learning models learn data representations within their hidden layer at multiple levels of abstraction (LeCun & al., 2015). We have constructed an LSTM model with five hidden layers of 128 neurons, with a batch size of 256 and an input layer of 14*128 (see Figure 4). The Network output layer has four nodes representing the amusement level (low, medium, high, very high).

```
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
lstm (LSTM)                  (None, 14, 128)           131584
_________________________________________________________________
lstm_1 (LSTM)                (None, 14, 128)           131584
_________________________________________________________________
lstm_2 (LSTM)                (None, 14, 128)           131584
_________________________________________________________________
lstm_3 (LSTM)                (None, 14, 128)           131584
_________________________________________________________________
lstm_4 (LSTM)                (None, 14, 128)           131584
_________________________________________________________________
lstm_5 (LSTM)                (None, 128)               131584
_________________________________________________________________
dropout (Dropout)            (None, 128)               0
_________________________________________________________________
dense (Dense)                (None, 4)                 516
=================================================================
```

Figure 4. – Neural network summary

To make sure our model can generalize on new content, we removed one video from the training set for all our participants and used this unseen video as our final validation test for the generated model. While keeping in mind that each model takes a large amount of time to train, we repeated our model training with 16 random videos (4 videos for each amusement intensity) to ensure our validation accuracy is unbiased by the chosen video and discussed the results below.

Furthermore, to ensure the model is usable on an unseen participant, we used a leave-one-group-out technique during the training and testing phase (Figure 5). More precisely, we trained the algorithm on all 8 participants and tested it on the last one not previously seen. We repeat this procedure so that each participant is used as the test set once. The mean accuracy of all algorithms can describe the algorithms' performance.

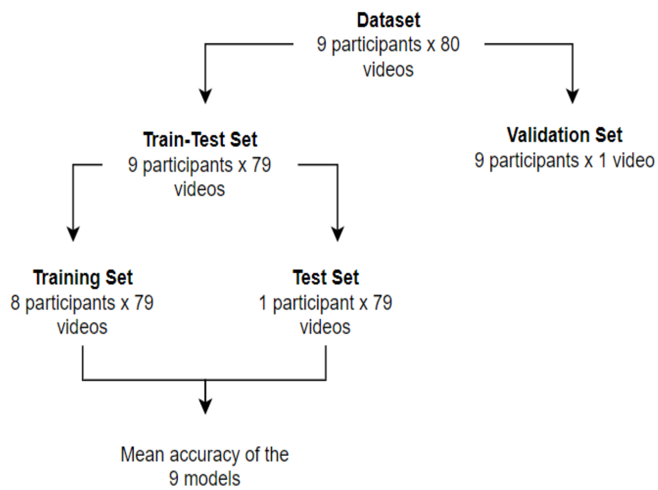

Figure 5. – Leave-One-Group-Out cross-validation method

Initially, we set the number of training epochs to 100 by cross-validation. We used early stopping techniques to prevent our model from overfitting (Figure 6). We set model monitoring on validation loss metrics during training. Early stopping was used to evaluate different learning rate values for the model. The weights of the best model were recorded with minimum validation loss.

```
earlyStopping = EarlyStopping(monitor='val_loss',
patience=10, verbose=0, mode='min')

mcp_save =
ModelCheckpoint('SavedModels/mdl_clf_wts.hdf5',
save_best_only=True, monitor='val_loss',
mode='min')

reduce_lr_loss =
ReduceLROnPlateau(monitor='val_loss', factor=0.1,
patience=7, verbose=1, epsilon=1e-4, mode='min')
```

Figure 6. – Early stopping code snapshot

Results and Discussion

Base Model

In the first place, we used the technique described above (figure 5) with the different validation sets (16 different videos) to create our base model, *i.e.* a model without any tunning. This allows us to better understand how our model is working and tune our pipeline in consequence. Any changes made to the model training will then be compared to this base model.

Generalization of the model

The validation accuracy of the base model, when tested on an unseen video, can be found in Table 1 under *validation accuracy*. Taken together, our models predict the amusement intensity of an unseen video with an average of 64.2% (std=14.7%) with a maximum accuracy of 88.9% (model #1) and a minimum of 32.9% (model #6). Since there is high variability in the validation accuracy, we cannot conclude that this specific algorithm can yet be transferable to other content. On the other hand, when the algorithm is tested on an unseen participant during the training phase, accuracy is more stable. The column *Mean Accuracy Training (STD)* of Table 1 shows the model's mean accuracy when tested on each of the unseen participants (n=9). We obtain a mean accuracy of 74.9% (std=3.8%) with an accuracy as high as 87.5% (model #1) and as low as 71.5% (model #9). This high accuracy and low standard deviation show that our model can predict the amusement level based on a participant's brain activity that it has never seen before.

While looking at each model's confusion matrix, we saw that the fourth class, namely very-high amusement, is well represented in none of the models (see example in Figure 7), which may cause the high variability observed in the validation accuracy.

Furthermore, the good results obtained could be due to overfitting. While looking at the data, we can see that when a video is funnier, the participant pressed the space bar (i.e. the active task used to define the time of interest) later in the video (correlation= -0.082, p= 0.013). This could lead to unbalanced classes.

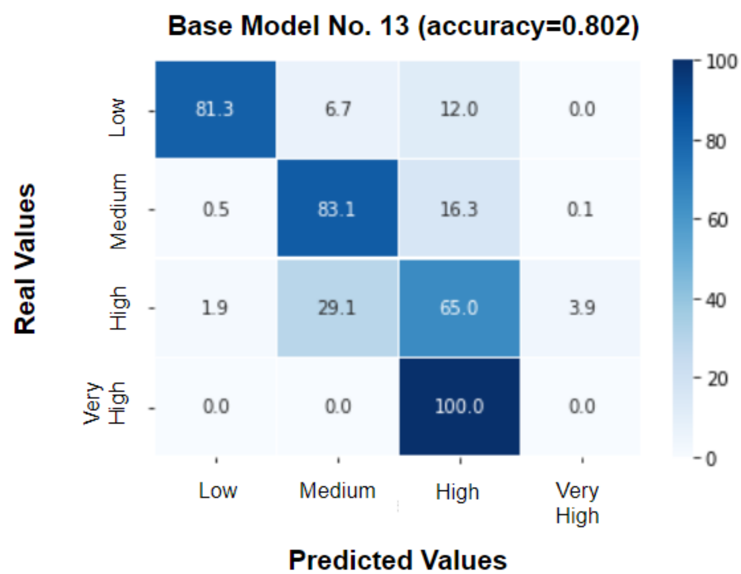

Figure 7. – Example of a confusion matrix during validation phase where very high amusement is unwell represented

Model with Class Weight

Weight Classes

To overcome the fact that our classes are unbalanced, we assign each class a weight during the training phase. We used an automatic function that looks at the distribution of labels and produces weights to equally penalize under or over-represented classes in the training set. Higher weights indicate that this class is underrepresented in our dataset. While each training model had different weights, the weight was very similar between models. A mean weight of 1.339

(std=0.014) was assign to low amusement, (std=0.009) to medium amusement, 0.591 (std=0.004) to high amusement and 6.284 (std=0.298) to very high amusement.

Generalization of the model

Out of the 16 models we trained, 10 of them saw their validation accuracy drop after the model was adjusted with weights (Table 1 under Validation Accuracy Gain/Loss). It confirms that our base model was overfitting on most of the models. We obtained a mean validation accuracy of 57.3% (std=13.7%). Like our base model, this high variability across the accuracy confirms that our model is not yet able to transfer perfectly to unseen videos. Our best model obtains an accuracy of 82.4%, and our lowest is at 31.1%.

Training and testing accuracy with the adjusted weights also dropped for all models. During the training and testing phase. We obtained a mean accuracy of 63.1% (std=3.0%) where our best model had an accuracy of 72.6% (std=13.3%) and our worst model had an accuracy of 59.6% (std=15.5%). The mean accuracy and every sub-model tested on unseen participants are above the theoretical chance level for four classes (chance=25%).

Having a low variability during the training phase helped us understand that the model's accuracy is not impacted by which participant the model is training and testing on. Since it's able to predict the amusement of an unseen participant, we can consider that a final version of this model should be transferable across unseen participants' brain activity.

Model Prediction

Across all the models created during this study, we were able to reach a validation accuracy as high as 82.41% (model #13 with the balanced group), which shows great promise to use brain activity as a predictor of amusement. Our pipeline also aims to ensure good transferability across unseen videos (low variability during validation) and unseen participants (low variability during testing). The high accuracy of our best model suggests that brain activity collected with a commercial headset could be used, with more research, to predict amusement.

While our model reaches a high accuracy level, we can see from the confusion matrix (Figure 8) that our model still has difficulty distinguishing between high and very high amusement. Our model can accurately predict the low and medium levels of amusement, but high and very high amusement are still inadequately predicted. It is possible that the model cannot classify between high and very high because the brain activity is more alike in those two categories than in low and medium amusement. Further research in this direction should include a comparison with another physiological signal such as facial expressions to confirm if the intensity of the smile is also different in this higher level of amusement. Moreover, inspired by Liu (2017), we believe that this problem could be improved by first creating a model that classifies the data between three types of funniness: low, medium and high (where high is a combination of high and very high amusement). This would be followed by a second model trained to classify especially between high and very high amusement, thus increasing our model prediction.

	Base Model		Model with Weights		
Validation Video	**Validation Accuracy**	**Mean Accuracy Training (STD)**	**Validation Accuracy**	**Validation Accuracy Gain/Loss**	**Mean Accuracy Training (STD)**
1	0.3602	0.7336 (0.143)	0.3115	-0.0487	0.634 (0.152)
2	0.5584	0.7489 (0.140)	0.5121	-0.0463	0.611 (0.157)
3	0.5339	**0.7912 (0.135)**	0.4899	-0.0440	0.602 (0.154)
4	0.5588	0.7420 (0.133)	0.6110	0.0522	0.616 (0.154)
5	**0.7443**	0.7264 (0.133)	0.6018	-0.1425	**0.726 (0.133)**
6	0.3289	0.7328 (0.139)	0.3537	0.0248	0.629 (0.159)
7	0.5169	0.7466 (0.132)	0.5560	0.0391	0.622 (0.149)
8	0.5086	0.7229 (0.126)	0.5328	0.0242	0.620 (0.152)
9	0.4874	0.7155 (0.136)	0.3923	-0.0951	0.596 (0.155)
10	0.6858	0.7286 (0.131)	0.6344	-0.0514	0.643 (0.137)
11	**0.7596**	0.7372 (0.135)	0.6795	-0.0801	0.606 (0.155)
12	**0.7401**	0.7339 (0.134)	0.6450	-0.0951	0.645 (0.150)
13	**0.8015**	0.7290 (0.133)	**0.8241**	0.0226	0.632 (0.140)
14	0.6978	**0.7577 (0.137)**	**0.7326**	0.0348	0.634 (0.144)
15	**0.7840**	**0.7714 (0.149)**	0.6228	-0.1612	0.616 (0.153)
16	0.6786	0.7251 (0.134)	0.6687	-0.0099	0.667 (0.161)

Tableau 1. – Model Generalization

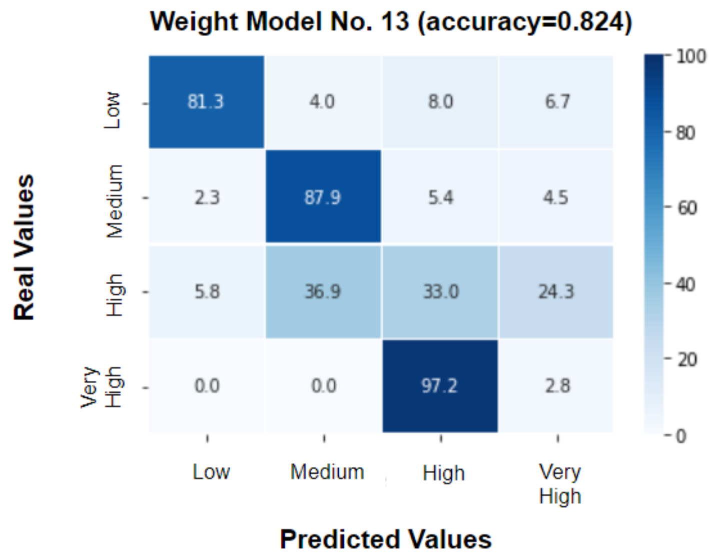

Figure 8. – Confusion matrix of the best weighted model during the validation phase

Conclusion

In this study, we aimed to develop an algorithm that predicts amusement based on EEG data from a commercial headset. Our main objective was to 1) find a model that can predict amusement with high accuracy and 2) develop a pipeline that can help us confirm that our model is transferable across both unseen participant and unseen content. Using an LSTM algorithm, we were able to obtain a model that can predict amusement with an accuracy of 82.4%. This high accuracy confirms that brain activity can accurately predict the amusement experienced by the subject.

Finally, the methodology developed and used in this study helped us understand and define, during the creation of the models, how our model could perform on an unseen participant and videos. In this study, our model had, on average, a low variability when testing on unseen participants, which supports that this model could predict the amusement of a new participant without having seen his brain activity. Unfortunately, our models tested on unseen videos were more variable. This lets us believe that we can still improve our model.

Classification of amusement in four-level (low, medium, high and very high) is our first step into creating an algorithm that can predict amusement. In this study, we confirm both the use of EEG data and LSTM to predict amusement.

Further Studies

Next, we want to improve our classification model by first creating a model that classifies the data between three types of funniness: low, medium, and high (where high is a combination of high and very high amusement). A second model would then be trained to classify between high and very high amusement. Furthermore, using the same pipeline, we will train an LSTM algorithm to predict a value between 0 (not funny) and 1 (very funny) on a continuous scale.

Acknowledgments

This project was possible thanks to the company *Beam Me Up*.

References

Alhagry, S., Fahmy, A. A., & El-Khoribi, R. A. (2017). Emotion recognition based on EEG using LSTM recurrent neural network. Emotion, 8(10), 355-358.

Aydın, S. (2019). Deep Learning Classification of Neuro-Emotional Phase Domain Complexity Levels Induced by Affective Video Film Clips. IEEE Journal of Biomedical and Health Informatics, 24(6), 1695-1702.

Barral, O., Kosunen, I., & Jacucci, G. (2017). No need to laugh out loud: Predicting humor appraisal of comic strips based on physiological signals in a realistic environment. ACM Transactions on Computer-Human Interaction (TOCHI), 24(6), 1-29.

Craik, A., He, Y., & Contreras-Vidal, J. L. (2019). Deep learning for electroencephalogram (EEG) classification tasks: a review. Journal of neural engineering, 16(3), 031001.

Duan, R. N., Zhu, J. Y., & Lu, B. L. (2013, November). Differential entropy feature for EEG-based emotion classification. In 2013 6th International IEEE/EMBS Conference on Neural Engineering (NER) (pp. 81-84). IEEE.

Gelkopf, M., Kreitler, S., & Sigal, M. (1993). Laughter in a psychiatric ward. Somatic, emotional, social and clinical influences on schizophrenic patients. Journal of Nervous and Mental Disease, 181(5), 283–289.

Herring, D. R., Burleson, M. H., Roberts, N. A., & Devine, M. J. (2011). Coherent with laughter: Subjective experience, behavior, and physiological responses during amusement and joy. International Journal of Psychophysiology, 79(2), 211-218.

Hirosaki, M., Ohira, T., Kajiura, M., Kiyama, M., Kitamura, A., Sato, S., & Iso, H. (2013). Effects of a laughter and exercise program on physiological and psychological health among community-dwelling elderly in Japan: Randomized controlled trial. Geriatrics & gerontology international, 13(1), 152-160.

Kim, J. H., Kim, B. G., Roy, P. P., & Jeong, D. M. (2019). Efficient facial expression recognition algorithm based on hierarchical deep neural network structure. IEEE Access, 7, 41273-41285.

LeCun, Y., Y. Bengio, and G. Hinton, Deep learning. Nature, 2015. 521(7553): p. 436-444.

Liu, Y. J., Yu, M., Zhao, G., Song, J., Ge, Y., & Shi, Y. (2017). Real-time movie-induced discrete emotion recognition from EEG signals. IEEE Transactions on Affective Computing, 9(4), 550-562.

Martin, R.A. (2007). The Social Psychology of Humor. In The Psychology of Humor, (Elsevier), pp. 113–152.

Peirce, J. W., Gray, J. R., Simpson, S., MacAskill, M. R., Höchenberger, R., Sogo, H., Kastman, E., Lindeløv, J. (2019). PsychoPy2: experiments in behavior made easy. Behavior Research Methods. 10.3758/s13428-018-01193

Savage, B. M., Lujan, H. L., Thipparthi, R. R., & DiCarlo, S. E. (2017). Humor, laughter, learning, and health! A brief review. Advances in Physiology Education.

Stöckli, S.; Schulte-Mecklenbeck, M.; Borer, S. & Samson, A.C. (2018) Facial expression analysis with AFFDEX and FACET: A validation study. Behavior Research Methods, 50 (4), 1446-1460.

Toupin, G., Saive, AL., Jerbi, K., (2021) Decoding the intensity of humour appreciation over time. In preparation for *Physiology & Behavior*.

Vrticka, P., Black, J. M., & Reiss, A. L. (2013). The neural basis of humour processing. Nature Reviews Neuroscience, 14(12), 860.

Watson, D., Clark, L. A., & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: The PANAS scales. Journal of Personality and Social Psychology, 54, 1063-1070.

Zhao, J., Yin, H., Wang, X., Zhang, G., Jia, Y., Shang, B., ... & Chen, L. (2020). Effect of humour intervention programme on depression, anxiety, subjective well-being, cognitive function and sleep quality in Chinese nursing home residents. Journal of Advanced Nursing, 76(10), 2709-2718.

Chapitre 4 – Discussion & Conclusion

Dans ce mémoire, nous avons exploré trois approches pour décoder de manière automatique et objective à quel point un participant perçoit le contenu humoristique d'une vidéo comme étant amusant. Des techniques d'intelligence artificielle, incluant l'apprentissage machine et l'apprentissage profond, nous ont permis de prédire l'appréciation humoristique des participants grâce aux caractéristiques des vidéos, du sourire des participants et de l'activité cérébrale mesurée par un casque EEG commercial.

Les données physiologiques comme les expressions faciales et l'activité cérébrale prédisent mieux l'intensité de l'amusement que les caractéristiques des vidéos, et ce, en plus d'avoir une meilleure granularité temporelle. Ces deux techniques ont leurs avantages et désavantages quant à leur implémentation dans de futures études. En effet, les expressions faciales sont facilement et rapidement implémentables dans de nouvelles études via l'ajout d'une webcam lors du visionnement du contenu humoristique. Plusieurs logiciels payants (ex: iMotions) et open-source (ex : OpenFace) permettent d'extraire les expressions faciales. Toutefois, la détection d'expressions faciales requiert d'être face à la caméra ce qui n'est pas toujours possible. Contrairement aux expressions faciales qui peuvent être imitées ou accentuées volontairement, l'activité cérébrale peut difficilement être contrôlée. De plus, l'émergence de nouvelles technologies en matière d'enregistrement cérébrale (ex: Emotiv, OpenBCI, etc.) rend l'utilisation de l'EEG de plus en plus accessible. Plusieurs logiciels (ex: Emotiv) ont implémenté la détection d'émotions grâce à l'EEG, par contre, il faut noter que les émotions positives tel que l'amusement ne sont généralement pas bien représentées et requièrent un outil personnalisé.

En plus d'avoir exploré différentes mesures objectives de l'appréciation de l'humour, le présent projet de maîtrise a parmi de mieux caractériser la dynamique temporelle et incrémentielle de l'appréciation de l'humour.

Retour sur l'étude 1 : Prédiction Physiologique et Comportementale

Résumé & Forces de l'étude

Lors de la première étude (chapitre 2), les participants ont évalué des vidéos plus ou moins drôles selon trois échelles (éveil, valence et drôlerie) pendant que leurs expressions faciales étaient enregistrées. Pour chacune des vidéos, nous avons extrait le mouvement, la saillance, la distance sémantique et la distance normalisée de Google. Dans un premier temps, nous avons utilisé ces caractéristiques ainsi que le sourire du participant afin de prédire l'appréciation humoristique des vidéos (de pas drôle à très drôle). Le sourire s'est avéré être le meilleur prédicteur de l'intensité de l'amusement tandis que les caractéristiques des vidéos ne permettaient que de distinguer le type de vidéo qui était visionnée (neutre versus humoristique). D'un point de vue fondamental, nous avons étudié la dynamique temporelle et incrémentielle de l'appréciation de l'humour. De manière consistante avec la nature complexe de l'humour et le fait que seule la "chute" évoque l'amusement, nous avons pu observer des performances de prédiction qui augmentent avec le temps, pour atteindre son maximum à la fin de la vidéo, une fois le dénouement visionné par les participants. De plus, nous avons pu définir qu'il existe un élément persistant et incrémentiel de l'amusement où plus la vidéo précédente est drôle, plus la suivante sera perçue comme drôle également.

Les forces de cette étude se trouvent dans 1) la validation des vidéos utilisées (pré-tests nombreux sur un grand échantillon de personnes assurant que chaque participant trouve plusieurs vidéos drôle, 2) la méthodologie innovante basée autant sur les caractéristiques visuelles des vidéos que la physiologie des participants, et 3) la prédiction au cours du temps du niveau d'amusement qui a permis de caractériser pour la première fois l'aspect temporel de l'humour.

Limites et Études Futures

Lors de cette étude, nous n'avons pas pris en considération les caractéristiques individuelles des participants telles que le genre et l'éducation. Ces différences individuelles pourraient expliquer pourquoi les caractéristiques de la vidéo n'ont pas aidé à décoder l'intensité de l'amusement. En effet, différentes personnes pourraient apprécier un contenu différent : une personne pourrait davantage préférer une personne qui tombe (beaucoup de mouvement) et une autre préférer des vidéos d'animaux qui font des faces bizarres (peu de mouvement). Nous prévoyons de futures analyses afin de comparer comment les caractéristiques démographiques influencent le sens de l'humour des participants ainsi que leurs évaluations émotionnelles des stimuli humoristiques. De plus, intégrer la dimension individuelle dans les analyses de cette étude permettrait de définir si certains groupes de gens (ex: grand sens de l'humour, préférence pour l'humour agressif, etc.) et certaines caractéristiques démographiques (ex: genre, âge, etc.) influencent les préférences pour certaines caractéristiques de la vidéo (ex: mouvement rapide). Pouvoir intégrer les caractéristiques individuelles avec les caractéristiques de la vidéo dans des analyses similaire permettrait de mieux comprendre comment l'un s'agence avec l'autre. Éventuellement, cette technique pourrait être utilisée pour prédire l'amusement général d'un participant lorsqu'un stimulus unique est utilisé.

L'utilisation de l'outil de détection de tags de Google, bien qu'elle soit automatique et facile à implémenter pour de futures études, est encore limitée par la qualité de l'algorithme. Afin de mieux comprendre l'importance de chacun des tags dans l'humour, une version des tags annotée manuellement par plusieurs participants serait plus représentative de la réalité.

Malheureusement, les vidéos étaient très différentes en termes de contenus entre les vidéos neutres (ex: animaux qui marchent, train en route, etc.) et drôles (ex: chat qui saute, enfants qui tombent, etc.) ce qui peut expliquer la facilité de l'algorithme à différencier entre les vidéos humoristiques et neutres, mais pas les différents niveaux d'humour. De futures études dans cette direction bénéficieraient grandement d'une plus grande diversité de contenu médiatique si différents types de vidéos sont employés.

Finalement, comparer un plus grand nombre de techniques entre-elles serait bénéfique. L'ajout de l'activité EEG, des battements cardiaques et de la conductance de la peau serait particulièrement intéressant.

Retour sur l’étude 2 : Prédiction grâce à l’activité cérébrale (EEG)

Résumé & Forces de l’étude

Lors de cette seconde étude, nous avons tenté de prédire le niveau d’amusement grâce à l’activité cérébrale brute mesurée par un casque EEG commercial de type Emotiv (14 électrodes). Un réseau d’apprentissage profond de type *Long-Short-Term-Memory* a été entraîné à prédire quatre intensités d’amusement (faible, moyen, élevé et très élevé) grâce à des segments de 1 seconde d’activité cérébrale.

Nous avons réussi à prédire, en moyenne, avec 63% d’exactitude (max=82% d’exactitude) le niveau d’appréciation humoristique que le participant était en train de visionner, montrant que l’activité cérébrale est une bonne option pour mesurer l'appréciation de l’humour. Le pipeline utilisé pour entraîner le modèle permet de comparer la qualité tant au niveau du transfert entre les participants (*capacité de l’algorithme à généraliser son apprentissage à de nouveaux participants qu’il n’avait pas encore vu)* qu’au travers du contenu (*capacité de l’algorithme à généraliser son apprentissage à de nouvelles vidéos*) ce qui rend le modèle plus facile à adapté pendant la période de conception selon les résultats.

Finalement, il est important de souligner que la méthodologie dans cet article suit une rigueur méthodologique le plus souvent retrouvée en recherche fondamentale et l’applique à un projet complété pour l’entreprise privée Beam Me Up, en partenariat avec Pr. Claude Frasson du DIRO, où différentes normes sont utilisées. Cette méthodologie apporte un regard innovateur et rigoureux quant à la conception des algorithmes pour l’entreprise.

Limites et Études Futures

Malgré le haut niveau d’exactitude du modèle entraîné, plusieurs aspects de cette étude pourraient être améliorés. Premièrement, le LSTM est un algorithme d’apprentissage profond récemment employé dans la détection d’émotions et peu d’études se sont intéressées spécifiquement à la prédiction entourant l’humour, le tout rendant difficile la comparaison de nos résultats avec d’autres déjà existants. Ainsi, de futures comparaisons avec des modèles plus communs comme le *Support Vector Machine* serait à prévoir afin d’avoir un seuil de comparaison pour les résultats actuels. Évidemment, l'amusement étant très subjectif, le faible échantillon de 10 participants limite la généralisation des résultats. En effet, un plus grand échantillon permettrait d'éliminer certains biais liés à l'état émotionnel vécu pendant l'expérience, les différents traits de personnalités et les caractéristiques démographiques.

Tel qu'observé dans la première étude comportementale (chapitre 2), le sourire corrèle fortement avec l'appréciation de l'humour. Ainsi, afin de poursuivre le développement d'une méthodologie robuste pour valider notre modèle, nous aimerions comparer les résultats de l'algorithme avec la présence de sourire. Plus précisément, nous aimerions observer la corrélation entre la prédiction faite par notre modèle et le sourire du participant lors de la même période. Si notre modèle performe correctement, la corrélation devrait être forte également.

À plus long terme, nous aimerions également travailler sur un modèle de régression permettant de prédire de manière continue le niveau d'amusement. Il est notamment possible de le faire tout en gardant l'algorithme LSTM utilisé lors du présent article. Il serait particulièrement intéressant de pouvoir observer le niveau de prédiction lorsque la vidéo est annotée comme drôle ou très drôle afin de voir si le niveau d'exactitude est plus faible tel qu'observé dans notre article.

Discussion Générale : Précision, Faisabilité & Intégration

L'objectif du mémoire était d'explorer différentes techniques de mesure objective de l'amusement. Plus précisément, nous avons exploré trois techniques différentes afin de quantifier objectivement l'état d'amusement à la suite d'un stimulus humoristique : les caractéristiques visuelles des vidéos (chapitre 2), les expressions faciales (chapitre 2) et l'activité cérébrale telle qu'enregistré par un casque commercial (chapitre 3).

Afin de pouvoir caractériser objectivement l'amusement et l'implémenter dans de futures études, ces trois techniques sont discutées dans cette section en termes de **1)** précision et biais, soient leur capacité à prédire l'amusement correctement ainsi que **2)** leur faisabilité et facilité d'intégration à de futures études.

Précision et biais des outils

Lors des études du présent mémoire, les techniques utilisant les composantes physiologiques, soit l'électroencéphalographie et les expressions faciales, semblent les plus précises pour prédire le niveau d'amusement. En effet, ces deux méthodes ont contribué activement à la prédiction de l'évaluation de l'amusement par les participants.

Les expressions faciales étaient fortement corrélées avec l'appréciation du contenu humoristique et ont permis un découpage temporel précis de l'amusement. L'utilisation du EEG a également permis d'atteindre jusqu'à 82% d'exactitude lors de la prédiction du niveau d'amusement. Il est important de noter que ces deux techniques ont permis une très grande précision temporelle puisque, contrairement à la valeur unique obtenue via l'évaluation de l'amusement (évaluer sur une échelle de 1-100 l'amusement vécu), elles ont permis d'enregistrer des données en continues pendant la tâche et lors du visionnement des vidéos. De plus, contrairement aux caractéristiques des vidéos, l'utilisation des caractéristiques physiologiques n'est pas dépendante du contenu de la vidéo, ce qui les rend d'autant plus précises et spécifiques au sentiment vécu par un individu à un moment spécifique.

Afin que l'implémentation des caractéristiques des vidéos soit plus précise, un modèle couplant les caractéristiques des vidéos avec des caractéristiques démographiques ou un des questionnaires sur les sens de l'humour tel que celui de Martin (2003) ou Svebak (1996, 2010) pourrai être développé. Cette composante ne serait toujours pas aussi précise et individuelle que l'utilisation d'une mesure physiologique puisqu'elle reposerait sur une généralisation du sens de l'humour, mais permettrait toutefois de pouvoir moduler certains résultats.

Faisabilité & Intégration lors d'études futures

Plus que de seulement regarder la précision de l'outil, il est important de considérer la facilité de leur implémentation dans de futures études. En effet, une mesure très précise n'est pas pour autant facile à implémenter lors d'une tâche en laboratoire ou lors de situations réelles.

Parmi les techniques utilisées lors du présent mémoire, on observe différentes forces et faiblesses au niveau de l'intégration de l'outil dans la tâche. Notamment, les expressions faciales ont été particulièrement simples à intégrer aux tâches expérimentales. La collecte des expressions faciales a pu être réalisée grâce à une caméra intégrée à un ordinateur portable (et webcam externe lorsque nécessaire), ce qui est très commun et abordable aujourd'hui. De plus, un nombre considérable d'études ont déjà été complétées sur l'utilisation des marqueurs AU6 et AU12 vis-à-vis de l'amusement. Ces dernières peuvent être facilement intégrées à de futures recherches via des outils payants tels qu'iMotions ou utiliser manuellement grâce à des libraires libres de droits comme OpenFace2.

Le EEG commercial a également été très simple à mettre en place puisqu'il vient avec un logiciel permettant l'enregistrement et une extraction simple des données. De plus, certaines versions de ces logiciels permettent de prédire, grâce à des algorithmes déjà en place, les émotions primaires telles que la joie grâce à l'activité cérébrale. Contrairement aux expressions faciales, la détection de l'amusement via l'activité cérébrale est encore à approfondir et demanderait davantage de recherche.

Finalement, en ce qui a trait aux caractéristiques des vidéos, leur plus grand avantage vient du fait qu'elles peuvent être utilisées à posteriori d'une étude puisqu'elles ne requièrent aucun interactions avec le participant. Ainsi, les caractéristiques des vidéos peuvent expliquer et aider à la modulation des résultats suite à une étude qui n'a pas mesuré directement l'amusement du participant. Contrairement à l'activité physiologique, les différentes caractéristiques des vidéos sont plus complexes à extraire. De plus, il existe plusieurs techniques pour mesurer la même caractéristique (ex : mouvement).

Basée sur les technologies actuelles, l'utilisation des expressions faciales semble être la technique la plus optimale pour mesurer le niveau d'amusement lorsque la mesure comportementale (à quel point le contenu est drôle) n'est pas possible.

Conclusion

Ce travail nous a permis de tester trois techniques pour prédire l'intensité de l'amusement généré par des vidéos humoristiques : les caractéristiques de la vidéo, les expressions faciales du participant et l'activité cérébrale. Nos résultats démontrent l'importance d'intégrer une mesure d’appréciation de l’humour et de l’intensité émotionnelle qu’il a évoqué (*i.e.,* amusement) dans les recherches et études cliniques portant sur l’humour puisque ceux-ci sont très variables entre les participants. Parmi les attributs testés dans ce mémoire, les expressions faciales et l’activité cérébrale semblent les mieux placées pour détecter objectivement et en temps réel l’amusement.

De plus, les connaissances fondamentales faites quant à la dynamique temporelle du niveau d’amusement permettront aux futures études de sélectionner le moment le plus propice pour mesurer l’amusement et ainsi maximiser les effets positifs de l’humour dans le cadre de nouveaux protocoles de recherche.

Références bibliographiques

Amir, O., Biederman, I., Wang, Z., & Xu, X. (2013). Ha Ha! Versus Aha! A direct comparison of humor to nonhumorous insight for determining the neural correlates of mirth. Cerebral Cortex, 25(5), 1405-1413.

Aydın, S. (2019). Deep Learning Classification of Neuro-Emotional Phase Domain Complexity Levels Induced by Affective Video Film Clips. IEEE Journal of Biomedical and Health Informatics, 24(6), 1695-1702.

Baid, H., & Lambert, N. (2010). Enjoyable learning: the role of humour, games, and fun activities in nursing and midwifery education. Nurse Education Today, 30(6), 548-552.

Barral, O., Kosunen, I., & Jacucci, G. (2018). No need to laugh out loud: Predicting humor appraisal of comic strips based on physiological signals in a realistic environment. ACM Transactions on Computer-Human Interaction (TOCHI), 24(6), 40.

Barrick, A. L., Hutchinson, R. L., & Deckers, L. H. (1990). Humor, aggression, and aging. The Gerontologist, 30(5), 675-678.

Bartolo, A., Benuzzi, F., Nocetti, L., Baraldi, P., & Nichelli, P. (2006). Humor comprehension and appreciation: an FMRI study. Journal of cognitive neuroscience, 18(11), 1789-1798.

Behrouz, S., Mazlom, S. R., Kooshiar, H., Aghebati, N., Asgharipour, N., & Behnam, H. (2017). Investigating the effect of humor therapy on chronic pain in the elderly living in nursing homes in Mashhad, Iran. Evidence Based Care, 7(2), 27-36.

Bekinschtein, T.A., Davis, M.H., Rodd, J.M., and Owen, A.M. (2011). Why Clowns Taste Funny: The Relationship between Humor and Semantic Ambiguity. Journal of Neuroscience 31, 9665–9671.

Bergen, D. (2009). Gifted children's humor preferences, sense of humor, and comprehension of riddles. Humor, 22(4), 419-436.

Berk, L. S., Felten, D. L., Tan, S. A., Bittman, B. B., & Westengard, J. (2001). Modulation of neuroimmune parameters during the eustress of humor-associated mirthful laughter. Alternative therapies in health and medicine, 7(2), 62-76.

Berk, L. S., Tan, S. A., Fry, W. F., Napier, B. J., Lee, J. W., Hubbard, R. W., ... & Eby, W. C. (1989). Neuroendocrine and stress hormone changes during mirthful laughter. The American journal of the medical sciences, 298(6), 390-396.

Bischetti, L., Canal, P., & Bambini, V. (2021). Funny but aversive: a large-scale survey of the emotional response to COVID-19 humor in the Italian population during the lockdown. Lingua, 249, 102963.

Brody, L. R. (2000). The socialization of gender differences in emotional expression: Display rules, infant temperament, and differentiation. Gender and emotion: Social psychological perspectives, 2.

Campbell, D. W., Wallace, M. G., Modirrousta, M., Polimeni, J. O., McKeen, N. A., & Reiss, J. P. (2015). The neural basis of humour comprehension and humour appreciation: The roles of the temporoparietal junction and superior frontal gyrus. Neuropsychologia, 79, 10-20.

Carretero-Dios, H., & Ruch, W. (2010). Humor appreciation and sensation seeking: Invariance of findings across culture and assessment instrument? Humor: International Journal of Humor Research, 23(4), 427-445. doi:10.1515/HUMR.2010.020

Caruana, F., Avanzini, P., Gozzo, F., Francione, S., Cardinale, F., & Rizzolatti, G. (2015). Mirth and laughter elicited by electrical stimulation of the human anterior cingulate cortex. Cortex, 71, 323-331.

Chan, Y. C., Chou, T. L., Chen, H. C., & Liang, K. C. (2012). Segregating the comprehension and elaboration processing of verbal jokes: an fMRI study. Neuroimage, 61(4), 899-906.

Chandrasekaran, A., Vijayakumar, A. K., Antol, S., Bansal, M., Batra, D., Zitnick, C. L., & Parikh, D. (2016). We are humor beings: Understanding and predicting visual humor. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 4603-4612).

Christie, I. C., & Friedman, B. H. (2004). Autonomic specificity of discrete emotion and dimensions of affective space: A multivariate approach. International journal of psychophysiology, 51(2), 143-153.

Colom, G. G., Alcover, C. T., Sánchez-Curto, C., & Zárate-Osuna, J. (2011). Study of the effect of positive humour as a variable that reduces stress. Relationship of humour with personality and performance variables. Psychology in Spain, 15(1), 9-21.

Crawford, S. A., & Caltabiano, N. J. (2011). Promoting emotional well-being through the use of humour. The Journal of Positive Psychology, 6(3), 237-252.

Dean, R. A. K., & Major, J. E. (2008). From critical care to comfort care: the sustaining value of humour. Journal of clinical nursing, 17(8), 1088-1095.

Detenber, B. H., Simons, R. F., & Bennett Jr, G. G. (1998). Roll 'em!: The effects of picture motion on emotional responses. Journal of Broadcasting & Electronic Media, 42(1), 113-127.

Dols, J. M. F., & Russ ell, J. A. (Eds.). (2017). The science of facial expression. Oxford University Press.

Ehrlé, N., Espi, P., Labire, J., Loizeau, A., Menard, C., & Bakchine, S. (2020). Impairments of humour comprehension in multiple sclerosis. Multiple sclerosis and related disorders, 39, 101443.

Ekman, P. (1999). Basic emotions. Handbook of cognition and emotion, 98(45-60), 16.

Fan, S., Shen, Z., Jiang, M., Koenig, B. L., Xu, J., Kankanhalli, M. S., & Zhao, Q. (2018). Emotional attention: A study of image sentiment and visual attention. In Proceedings of the IEEE Conference on computer vision and pattern recognition (pp. 7521-7531).

Fang, Y., Lin, W., Chen, Z., Tsai, C. M., & Lin, C. W. (2013). A video saliency detection model in compressed domain. IEEE transactions on circuits and systems for video technology, 24(1), 27-38.

Fonzi, L., Matteucci, G., & Bersani, G. (2010). Laughter and depression: hypothesis of pathogenic and therapeutic correlation. Rivista di psichiatria, 45(1), 1-6.

Froehlich, E., Madipakkam, A. R., Craffonara, B., Bolte, C., Muth, A. K., & Park, S. Q. (2021). A short humorous intervention protects against subsequent psychological stress and attenuates cortisol levels without affecting attention. Scientific Reports, 11(1), 1-10.

Ghaffari, F., Dehghan-Nayeri, N., & Shali, M. (2015). Nurses’ experiences of humour in clinical settings. Medical Journal of the Islamic Republic of Iran, 29, 182.

Girard, J. M., Cohn, J. F., Yin, L., & Morency, L. P. (2020). Reconsidering the Duchenne Smile: Examining the Relationships between the Duchenne Marker, Smile Intensity, and Positive Emotion.

Goel, V., & Dolan, R. J. (2001). The functional anatomy of humor: segregating cognitive and affective components. Nature neuroscience, 4(3), 237.

Gonot-Schoupinsky, F. N., & Garip, G. (2018). Laughter and humour interventions for well-being in older adults: A systematic review and intervention classification. Complementary therapies in medicine, 38, 85-91.

Gozukara, E. (2016). Moderating role of gender in the relationship between humor styles and group cohesiveness. International Journal of Academic Research in Business and Social Sciences, 6(2), 123-135.

Henry, J. D., Phillips, L. H., Ruffman, T., & Bailey, P. E. (2013). A meta-analytic review of age differences in theory of mind. Psychology and Aging, 28(3), 826.

Herring, D. R., Burleson, M. H., Roberts, N. A., & Devine, M. J. (2011). Coherent with laughter: Subjective experience, behavior, and physiological responses during amusement and joy. International Journal of Psychophysiology, 79(2), 211-218.

Hofmann, J., Platt, T., & Ruch, W. (2017). Laughter and smiling in 16 positive emotions. IEEE Transactions on Affective Computing, 8(4), 495-507.

Iidaka, T. (2016). Humor appreciation involves parametric and synchronized activity in the medial prefrontal cortex and hippocampus. Cerebral Cortex, 27(12), 5579-5591.

Ko, H. J., & Youn, C. H. (2011). Effects of laughter therapy on depression, cognition and sleep among the community-dwelling elderly. Geriatrics & Gerontology International, 11(3), 267-274.

Kohn, N., Kellermann, T., Gur, R. C., Schneider, F., & Habel, U. (2011). Gender differences in the neural correlates of humor processing: implications for different processing modes. Neuropsychologia, 49(5), 888-897.

Krumhuber, E. G., & Manstead, A. S. (2009). Can Duchenne smiles be feigned? New evidence on felt and false smiles. Emotion, 9(6), 807.

Liu, Y. J., Yu, M., Zhao, G., Song, J., Ge, Y., & Shi, Y. (2017). Real-time movie-induced discrete emotion recognition from EEG signals. IEEE Transactions on Affective Computing, 9(4), 550-562.

Martin, R. A., & Ford, T. (2018). The psychology of humor: An integrative approach. Academic press.

Martin, R. A., Puhlik-Doris, P., Larsen, G., Gray, J., & Weir, K. (2003). Individual differences in uses of humor and their relation to psychological well-being: Development of the Humor Styles Questionnaire. Journal of research in personality, 37(1), 48-75.

Mathew, H. E., & Vijayalakshmi, V. (2017). Changing definitions of work and play: Importance of workplace humour. Psychological Studies, 62(1), 12-20.

Mathies, C., Chiew, T. M., & Kleinaltenkamp, M. (2016). The antecedents and consequences of humour for service: A review and directions for research. Journal of Service Theory and Practice.

McRae, K., Ochsner, K. N., Mauss, I. B., Gabrieli, J. J., & Gross, J. J. (2008). Gender differences in emotion regulation: An fMRI study of cognitive reappraisal. Group processes & intergroup relations, 11(2), 143-162.

Mobbs, D., Greicius, M. D., Abdel-Azim, E., Menon, V., & Reiss, A. L. (2003). Humor modulates the mesolimbic reward centers. Neuron, 40(5), 1041-1048.

Moran, C. C. (1996). Short-term mood change, perceived funniness, and the effect of humor stimuli. Behavioral medicine, 22(1), 32-38.

Moran, J. M., Wig, G. S., Adams Jr, R. B., Janata, P., & Kelley, W. M. (2004). Neural correlates of humor detection and appreciation. Neuroimage, 21(3), 1055-1060.

Murata, K. (2014). An empirical cross-cultural study of humour in business meetings in New Zealand and Japan. Journal of Pragmatics, 60, 251-265.

Pérez-Aranda, A., Hofmann, J., Feliu-Soler, A., Ramírez-Maestre, C., Andrés-Rodríguez, L., Ruch, W., & Luciano, J. V. (2019). Laughing away the pain: A narrative review of humour, sense of humour and pain. European Journal of Pain, 23(2), 220-233.

Rasheed, Z., Sheikh, Y., & Shah, M. (2005). On the use of computable features for film classification. IEEE Transactions on Circuits and Systems for Video Technology, 15(1), 52-64.

Shammi, P., & Stuss, D. T. (2003). The effects of normal aging on humor appreciation. Journal of the International Neuropsychological Society, 9(6), 855-863. 370.

Shatz, M. A., & LoSchiavo, F. M. (2006). Bringing life to online instruction with humor. Radical Pedagogy, 8(2), 8.

Simons, R. F., Detenber, B. H., Cuthbert, B. N., Schwartz, D. D., & Reiss, J. E. (2003). Attention to television: Alpha power and its relationship to image motion and emotional content. Media psychology, 5(3), 283-301.

Soleymani, M., Lichtenauer, J., Pun, T., & Pantic, M. (2011). A multimodal database for affect recognition and implicit tagging. IEEE transactions on affective computing, 3(1), 42-55.

Svebak, S. (1996). The development of the Sense of Humor Questionnaire: From SHQ to SHQ-6.

Svebak, S. (2010). The Sense of Humor Questionnaire: Conceptualization and review of 40 years of findings in empirical research. Europe’s journal of psychology, 6(3), 288-310.

Svebak, S., Martin, R. A., & Holmen, J. (2004). The prevalence of sense of humor in a large, unselected county population in Norway: Relations with age, sex, and some health indicators. Humor, 17(1/2), 121-134.

Tan, S. A., Tan, L. G., Lukman, S. T., & Berk, L. S. (2007). Humor, as an adjunct therapy in cardiac rehabilitation, attenuates catecholamines and myocardial infarction recurrence. Advances in mind-body medicine, 22(3-4), 8-12.

Thaler, A., Posen, J., Giladi, N., Manor, Y., Mayanz, C., Mirelman, A., & Gurevich, T. (2012). Appreciation of humor is decreased among patients with Parkinson’s disease. Parkinsonism & related disorders, 18(2), 144-148.

Thorson, J. A., Powell, F. C., Sarmany-Schuller, I., & Hampes, W. P. (1997). Psychological health and sense of humor. Journal of clinical psychology, 53(6), 605-619.

Uekermann, J., Channon, S., Lehmkämper, C., Abdel-Hamid, M., Vollmoeller, W., & Daum, I. (2008). Executive function, mentalizing and humor in major depression. Journal of the International Neuropsychological society, 14(1), 55-62.

Vilaythong, A. P., Arnau, R. C., Rosen, D. H., & Mascaro, N. (2003). Humor and hope: Can humor increase hope?. Humor, 16(1), 79-89.

Vrticka, P., Black, J. M., & Reiss, A. L. (2013). The neural basis of humour processing. Nature Reviews Neuroscience, 14(12), 860.

Wang, R. W., Kuo, H. C., & Chuang, S. W. (2017). Humor drawings evoked temporal and spectral EEG processes. Social cognitive and affective neuroscience, 12(8), 1359-1376.

Wang, S., & Ji, Q. (2015). Video affective content analysis: a survey of state-of-the-art methods. IEEE Transactions on Affective Computing, 6(4), 410-430.

Watson, K. K., Matthews, B. J., & Allman, J. M. (2006). Brain activation during sight gags and language-dependent humor. Cerebral cortex, 17(2), 314-324.

Willinger, U., Hergovich, A., Schmoeger, M., Deckert, M., Stoettner, S., Bunda, I., ... & Auff, E. (2017). Cognitive and emotional demands of black humour processing: the role of intelligence, aggressiveness and mood. Cognitive processing, 18(2), 159-167.

Wu, C. L., Lin, H. Y., & Chen, H. C. (2016). Gender differences in humour styles of young adolescents: Empathy as a mediator. Personality and Individual Differences, 99, 139-143.

Zhao, J., Yin, H., Zhang, G., Li, G., Shang, B., Wang, C., & Chen, L. (2019). A meta-analysis of randomized controlled trials of laughter and humour interventions on depression, anxiety and sleep quality in adults. Journal of advanced nursing, 75(11), 2435-2448.

Zheng, H., Chen, T., You, Q., & Luo, J. (2017, September). When saliency meets sentiment: Understanding how image content invokes emotion and sentiment. In 2017 IEEE International Conference on Image Processing (ICIP) (pp. 630-634). IEEE.

Annexes

Plusieurs sections du code utilisé lors du premier article se retrouvent sur le lien GitHub suivant. On y retrouve notamment : la tâche, l'extraction des caractéristiques des vidéos et le pipeline d'apprentissage automatique.

→ https://github.com/Rammen/MasterProject

Les extraits des vidéos humoristiques utilisés lors des deux expériences se retrouvent dans une collection publique sur Youtube.

→ https://youtube.com/playlist?list=PLcBTyKtg-JVDx9nAnzD8lnmIvXfH9avnL