Playable

INSIGHTS

Q1 2018

Computer Vision Analysis Of Marketing Videos

TABLE OF CONTENTS

INTRODUCTION

Playable upgrades email marketing to deliver high quality native autoplay video.

Marketers and publishers upload a source video to Playable then use our Smart Video Editor to slice and dice the content, and enhance the video with captions, a watermark, an end-card or footer to generate their final edit.

The Editor is smart because it has learnt, from all the data that Playable collects about video engagement, that these enhancements have a positive impact on audience engagement.

The Editor is also smart because it uses computer vision to help you create better videos that engage more with your audience.

There's a quick explainer of "computer vision" towards the end of this report, if it's an unfamiliar term. TL;DR: computer vision is the technology behind self-driving cars, and our Smart Video Editor!

Our Editor uses computer vision to identify what is shown in the video content, and to assist in the creation of your final edit.

This editing generates an interesting dataset that we explore in this edition of Playable Insights: **Computer Vision Analysis of Marketing Videos** – *what do we remove or keep when editing video?*

EXECUTIVE SUMMARY

1 trillion pixels.

That's what we started with, contained within the source material for a random collection of video marketing campaigns.

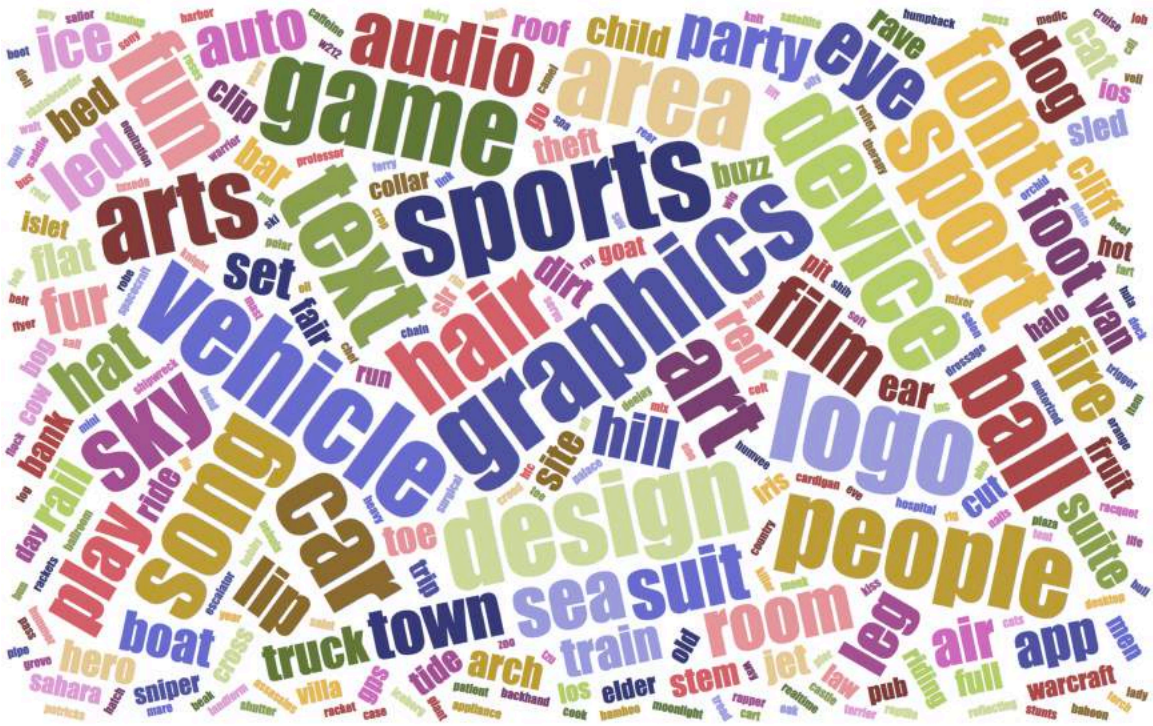
Marketers edited those source videos into their marketing videos, using Playable's Smart Video Editor.

The edits are 6 to 10 seconds duration, with no audio, created with the objective of engaging viewers to take action.

77% of the pixels were removed... Which ones?

When marketers edit videos, what is removed and what is kept?

Analyzed using computer vision, and visualized as a tag cloud, the videos went from containing these …



… to these:

The differences between the tag clouds above reveal valuable insights about video editing, and this report aims to highlight some of those.

We confirm that **Person** is the most prominent thing detected, in both source videos and edited marketing videos, claiming a massive **6% of all screentime**.

However, **People** – meaning many persons together – drops from being the 5th most prominent thing detected in source videos, to the 15th most prominent thing detected in edited marketing videos.

Using data from computer vision analysis of marketing videos, we reveal a selection of what marketers prefer to keep in their videos, including “toy”, “food” and “home”.

And a sample of what marketers prefer to remove, including “eyewear”, “sitting” and – controversially – “cats”.

Please continue reading to find out more. Or if you prefer to see with your own eyes how Playable has assembled this unique dataset, please visit **https://playable.video** where you can upload your own source video and use our Smart Video Editor to create your marketing video, including the ability to navigate your video by computer vision.

SOURCE VIDEOS

This table shows the labels detected in the source videos, ranked by prominence. See Methodology (page 11) for an explanation of how "prominence" is calculated. The labels are clustered, for ease of comparison. **Person** is the most prominent cluster:

Rank	Cluster	Related labels included in the cluster
1	**Person**	interaction, smile, facial expression, sitting, presentation, public speaking, singing, emotion, speech, happiness, pedestrian, cheering, standing, learning, student, orator, referee, professional, humour, applause, ...
2	**Sports**	ball game, team sport, physical fitness, physical exercise, soccer, basketball, indoor games and sports, motorsport, equestrianism, extreme sport, pool, cycling, boating, baseball, bat and ball games, athletics, …
3	**Location**	stage, music venue, arena, nightclub, tourist attraction, airport, …
4	**Structure**	stadium, arena, shooting range, race track, golf course, tennis court, …
5	**People**	crowd, audience, troop, family
6	**Vehicle**	car, aircraft, motorcycle, boat, bicycle, ship, watercraft, train, truck, …
7	**Artwork**	drawing, statue, anime, line art, painting, floral design, anime, mural, …
8	**Entertainment**	performance, arts, singing, performance art, film, theatre, …
9	**Event**	ceremony, meeting, interview, championship, combat, holiday, festival, …
10	**Communication**	conversation, …
11	**Graphics**	logo, clip art, text, …
12	**Geographical feature**	city, horizon, forest, wilderness, landscape, rural area, town, mountain, hill, grassland, landmark, residential area, meadow, pasture, coast, river, …
13	**Music**	song, singing, instrument, …
14	**Organization**	company, team, business, military, club, law enforcement, community, …
15	**Technology**	gadget, electronic device, led display, communication device, robot, …

Ranked next after these are Building, Computer, City, Plant, Software, Art, Clothing, Car, Airplane, and Animal.

An important distinction is made between **Person**, meaning a singular human, and **People**, meaning a plurality of people.

Note: In this report, it's possible for a label to be included in more than one cluster, e.g. "singing" is included with both Music and Person. Also it's possible that a label, e.g. "car", included in a cluster "vehicle", qualifies as its own cluster if it has enough prominence of its own and includes related labels, e.g. "sports car", "luxury vehicle".

EDITED MARKETING VIDEOS

Here we compare the prominence rank of label clusters from the source videos, shown on the left, with the rank of label clusters from the edited marketing videos, shown on the right:

	SOURCE VIDEOS
1	**Person**
2	**Sports**
3	**Location**
4	**Structure**
5	**People**
6	**Vehicle**
7	**Artwork**
8	**Entertainment**
9	**Event**
10	**Communication**
11	**Graphics**
12	**Geo. feature**
13	**Music**
14	**Organization**
15	**Technology**

	MARKETING VIDEOS
1	**Person**
2	**Sports**
3	**Artwork**
4	**Vehicle**
5	**Graphics**
6	**Location**
7	**Structure**
8	**Computer**
9	**Airplane**
10	**City**
11	**Building**
12	**Event**
13	**Organization**
14	**Technology**
15	**People**

Graphics (including logo, clip art, text) is a winner, meaning it's significantly more prominent in final edits than source videos. This reflects the need for short-form silent marketing videos to include a clear call to action such as captions or an end-card.

People (including crowd, audience, troop, family) is a loser, meaning that it's significantly less prominent in final edits than source videos. This could reflect the fact that interaction between people is hard to convey in a short-form silent marketing video.

Similarly, **Communication** drops off the chart from 10th to 97th. The marketers who are doing this editing are correct: Playable's video engagement data confirms that "mouth flapping" (talking heads with no sound) does not help your video engagement metrics!

.

WINNERS AND LOSERS

Digging into the detail, here we show individual labels from the Top 250 Most Prominent, and how their prominence changes from source videos to edited marketing videos.

Winners – What do marketers keep in their videos?

This is a selection of what marketers are choosing to keep in their video marketing, based on the objective of engaging their audiences:

Label	Rank in Source Videos	Rank in Marketing Videos
news	**246**	**80**
education	**215**	**64**
toy	**185**	**41**
classroom	**212**	**82**
food	**196**	**76**
air travel	**171**	**60**
skyscraper	**189**	**97**
learning	**146**	**55**
company	**154**	**67**
home	**147**	**68**

We leave the interpretation of these results as an exercise for the reader.

Losers – What do marketers remove from their videos?

And this is a selection of what marketers are choosing to remove from their video marketing, based on the objective of engaging their audiences:

Label	Rank in Source Videos	Rank in Marketing Videos
cheering	**66**	**217**
special effects	**75**	**214**
body of water	**110**	**235**
music venue	**59**	**183**
grass	**97**	**220**
concert	**52**	**173**
sitting	**37**	**150**
horse racing	**118**	**228**
emotion	**64**	**169**
eyewear	**58**	**160**

… and what about the cats?

No analysis of video content would be complete without looking at cat prominence.

You'll notice "cat" in the top right corner of the tag cloud on page 4 of this report. Nestled next to "dog", by sheer coincidence.

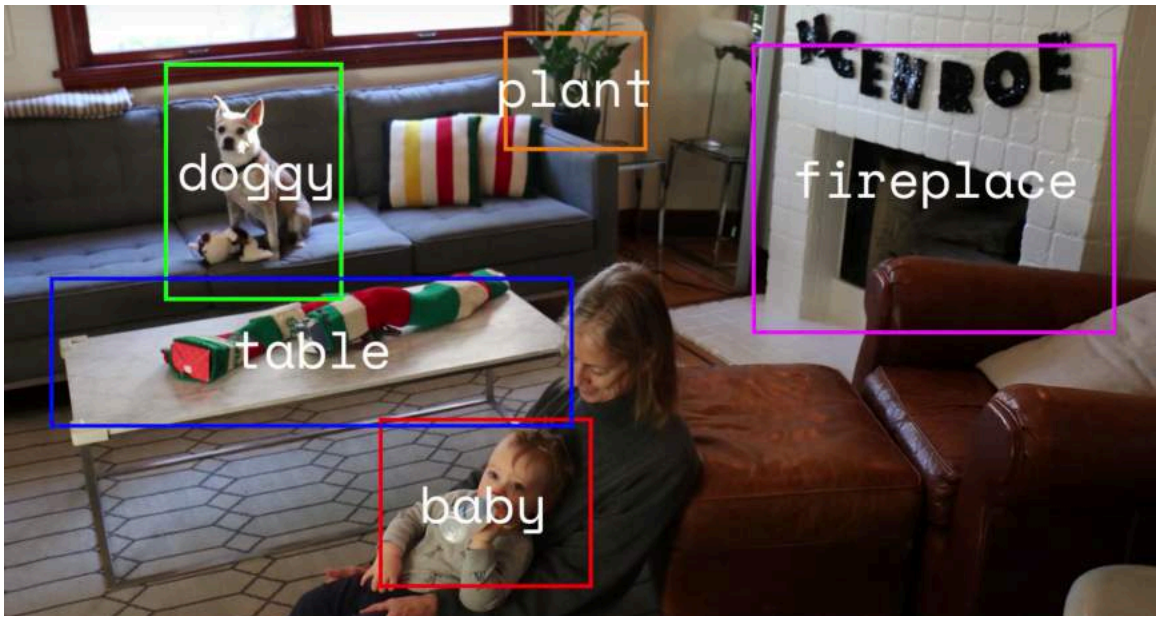
Cats rank as the 494th most prominent label in source videos, however after editing by humans, they drop in rank to 694th – proving empirically that, despite the great job that catkind has done to distribute its content across the interwebs, human marketers don't believe that cat video is useful to drive marketing results.

Or perhaps that's just what the cats want us to believe. Here are some other related results:

Label	Rank in Source Videos	Rank in Marketing Videos	Comment
cat	**494**	**694**	**See above**
dog	**236**	**364**	**Could do better**
person	**1**	**1**	**Clearly the dominant species**
computer	**17**	**8**	**One to watch!**

WHAT IS COMPUTER VISION?

Computer vision is teaching computers to solve this same problem you can solve with your own eyes:



Screenshot from https://youtu.be/eQLcDmfmGB0

A computer vision algorithm examines the pixels that comprise every frame of a video, and the differences between frames, to detect points, edges, corners, basic shapes and movements.

Higher-level objects can then be recognized, such as the dog in the above example, and the video is annotated with labels such as “dog”.

Self-driving cars detect the environment around them by using specialized computer vision technology.

General computer vision services are available from Google VideoIntel, IBM Watson and Amazon Rekognition.

At Playable we use Google and Amazon and our own technology.

METHODOLOGY

This report reflects anonymized and aggregate data from a random sample of 1,000 videos that were manually edited for audiences of over 10 million viewers, in over 100 countries during January and February of 2018.

1 million video frames, containing 1 trillion pixels, were analyzed using computer vision services powered by Google VideoIntel, to generate 15,000 labels to describe what's in the videos.

For the tag clouds in this report, the labels were sized by prominence, calculated as their duration onscreen in the videos, weighted by the confidence with which the computer vision algorithms recognize them. For example, a label that is onscreen for longer scores higher, and a label that has been recognized with low confidence scores lower.

Thanks to @jasondavies for the tag cloud visualizations.

For the tables in this report, and to help with the analysis, semantic clustering is used to group together similar labels. For example, "smile", "singing" and "standing" are clustered together with "person".

Thanks to Google for their Knowledge Graph which powered the semantic clustering.

THANK YOU

Thanks for your interest in Playable and video marketing and computer vision.

We welcome all your feedback on this **Playable Insights Q1 2018** report, and your requests for additional analysis in our future white papers.

Our next report coming in Q2 will focus on another interesting slice through the data.

Playable is headquartered in Silicon Valley, with offices in Sunnyvale and Sydney, and customer success operations in New York and Melbourne.

For more details, please contact:

Bob Hitching
CEO
Playable, Inc.
+1 650 289 8577
bob.hitching@playable.video