

## Shazam (2014)

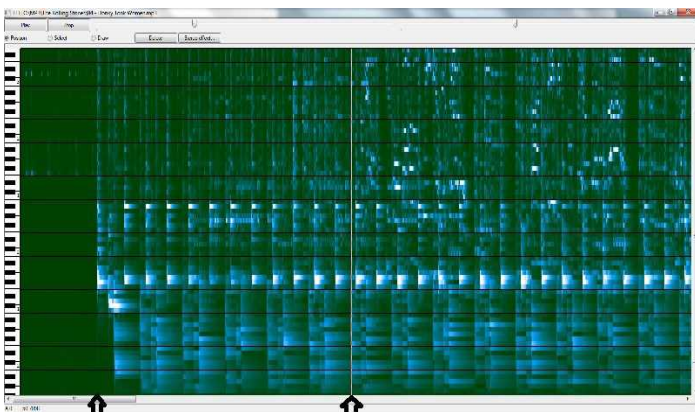
Comment Shazam peut-il reconnaître une musique ? C'est quand même assez bluffant.

### Alors, comment que ça marche ?

Shazam est capable d'identifier en quelques secondes la musique que vous êtes en train d'écouter même en environnement bruyant. On dirige le micro du téléphone vers un haut-parleur, on clique sur un bouton et quelques secondes après, le titre du morceau s'affiche, avec les liens qu'il faut pour l'acheter, évidemment. Ça ne marche pas en chantant la cousine Lison, c'était juste pour l'exemple, il faut que ce soit une musique enregistrée. (Pour les chants en live, il y a une autre appli qui s'appelle Midomi mais qui a un taux de réussite bien inférieur à Shazam et il faut chanter correctement.

Ce qui nous intéresse ici, c'est de savoir comment Shazam fait pour reconnaître quasi-instantanément 10 secondes de son parmi 8 millions de morceaux de musique de 3 minutes chacun, soit un total 45 ans de musique en continue, et ceci même lorsque la musique est diffusée par une mauvaise sono et captée par un mauvais micro dans un environnement bruyant.

Première chose, la technologie mise en œuvre par Shazam n'est pas la propriété de la firme à la pomme, c'est la propriété de la startup Landmark Digital fondée par deux mathématiciens.



(Diagramme fait avec un petit programme gratuit qui s'appelle FFT)

La figure précédente c'est le spectrogramme du début de la chanson « Honky tonk woman » des Rolling Stones. Il s'agit d'une représentation temps/fréquences : en abscisse on a le temps, et en ordonné on a les fréquences présentes dans la chanson à un instant donné.

L'intensité du point représente la « puissance » de la fréquence dans la chanson à l'instant donné, ils auraient pu faire une représentation 3D mais ça aurait été plus complexe.

On voit bien la structure du morceau, au début Charlie Watts tape sur une casserole pendant quelques secondes (jusqu'à la première flèche), il n'y a que quelques petits points, ensuite on a la batterie et l'intro très spéciale à la guitare, et ensuite au niveau de la deuxième flèche Mick Jagger se met à chanter et là le diagramme devient plus complexe.

La technologie Shazam/Landmark s'appuie sur ces

diagrammes pour calculer des « empreintes » de chaque chanson qu'ils ont dans leur base de données, ils ont 8 millions de morceaux dans la base.

Shazam compare des "empreintes" calculées à des instants remarquables du morceau, par exemple lorsque des notes apparaissent nettement dans le spectrogramme du morceau.

Le brevet est extrêmement vague sur le problème clé du calcul des empreintes. Il mentionne une liste de caractéristiques du son qui pourraient être utilisées, et indique qu'un choix judicieux permet de s'affranchir de diverses perturbations du son, en particulier de décalages fréquentiels, mais surtout que l'empreinte peut être ramenée à un nombre de 32 bits seulement. La taille totale de la base Shazam est donc relativement modeste (une cinquantaine de Giga Octets)

Alors comment ça marche. Vous enregistrez 10 secondes du morceau que vous voulez reconnaître, l'appli Shazam de votre smartphone extrait le spectrogramme de l'échantillon (un peu comme mon programme FFT) et l'envoi au serveurs de Shazam/Landmark, le serveur calcul les empreintes de l'échantillon et les compare à celles qu'il a en base.

Pour chaque empreinte, Shazam regarde à quels morceaux elle pourrait appartenir en utilisant simplement l'empreinte comme une clé de recherche dans la base de données.

Une empreinte donnée se retrouve en moyenne dans 3 morceaux seulement (ça c'est une information donnée par Landmark, c'est une donnée constatée). Avec 10 secondes d'enregistrement, on tire environ une centaine d'empreintes, il faut donc comparer une centaine d'empreintes qui vont chacune correspondre à environ 3 morceaux, donc on se retrouvera avec environ 300 morceaux potentiellement candidats.

Le serveur Shazam fait alors une simplification (information Landmark) : si 2 empreintes issues de l'échantillon correspondent à un même morceau, on considère que c'est celui-là et on arrête la recherche à ce stade, ça marche presque à tous les coups.

Tout ça pour dire aussi que l'intelligence de cette appli de reconnaissance musicale est principalement située du côté des serveurs Landmark, ce qui se passe du côté du smartphone est finalement très basique.

La technologie de Shazam/Landmark permet aussi d'autres applications, comme reconnaître la musique utilisée dans les bandes son des vidéos sur YouTube et la virer si elle est soumise à copyright, c'est d'ailleurs très chiant quand on fait du montage vidéo et qu'on diffuse les vidéos sur Youtube, souvent la vidéo est rejetée à cause de la musique, maintenant je sais pourquoi.

Cette application est gratuite pour le moment. Landmark se rémunère avec la publicité et prend un pourcentage sur les ventes iTunes réalisées grâce à l'appli.

Une prise de tête signée JB mais largement inspirée du blog du Dr Goulu