Challenges of Music Recommendation Software

SERGEJ LUGOVIĆ
Polytechnic of Zagreb
Vrbik 8, 10 000 Zagreb
CROATIA
slugovic@tvz.hr

NIVES MIKELIĆ PRERADOVIĆ
Department of Information and Communication Sciences
Faculty of Humanities and Social Sciences, University of Zagreb
I. Lucica 3, 10000 Zagreb
CROATIA
nmikelic@ffzg.hr
http://www.ffzg.unizg.hr/

Abstract: In this paper, we present results of the analysis of a music identification application – Shazam Our aim was to discover if applications such as Shazam can improve organisational systems of performing rights organizations (PROs) in the area of precise monitoring of the songs used in the night club environment. The aim of the research was to evaluate how precise Shazam is in two different environments, the first one being a night club (where the music is loud and the signal to noise ratio is high) and the other one being home studio environment, where the signal to noise ratio is low. Each song was tagged for 10 seconds in two different environments and Shazam created two different audio fingerprints for each song based on some of the anchors of the simplified spectrogram and the target area between them. The conclusion is that Shazam application is not very accurate in the night club environment and, at this stage, cannot be used by performing rights organizations to monitor the music played in public spaces. In order for royalties to be collected and distributed correctly, one needs to have more precise instruments that offer the guarantee that music creators will get their due.

Key-Words: music recommendation software, music information retrieval, Shazam, Galaxy1, iPhone5, transaction cost theory, performing rights organizations

1 Introduction

The development of smart phones and wireless internet in the last decade significantly increased the research interest in the field of Music Information Retrieval (MIR), since users started to frequently use available technology to discover, share and buy music.

Also, along with the growing interest in MIR researcher, grew the number of mobile music identification application, which identify unknown songs users hear on the radio, link users to related services such as online music stores, concert ticket vendors and artist merchandise, offering users the possibility to interact with social networks sharing identified content (i.e. newly discovered music).

Apart from showing the user related song metadata such as artist name, song title, album name, lyrics and artist biographies, such applications claim to engage users more deeply with what they are listening to.

The current commercial market leader in the field of music identification is a company called Shazam, founded in 1999 by Chris Barton, Philip Inghelbrecht, Avery Wang and Dhiraj Mukherjee [1]. It became one of the top 10 downloaded applications for iPhone [2]. In the beginning of 2014, Shazam plans to attract more than 1 billion dollars through Initial Public Offering (IPO).

Shazam is a music identification application which gathers a brief sample of music being played (on a mobile phone or some other device), creates an acoustic fingerprint (a time-frequency graph called spectrogram) based on that sample and searches a central database of more than 11 million songs for a match.

If the retrieval result is positive, Shazam provides the user with information such as artist, album, title, genre, music label, lyrics, a thumbnail image of the song/album artwork, links to download the song on iTunes or the Amazon store and, where relevant, shows the song's video on YouTube and gives the option of playing the song on Spotify.

It is a free or low-cost application for Androids, Apple iPhones, BlackBerrys, Nokia smart phones, Windows Phone devices and Sony Ericsson phones.

The application can identify pre-recorded music being broadcast from any source, such as a radio, television, cinema or club, provided that the background noise level is not high enough to prevent an acoustic fingerprint being taken, and that the song is present in its central database.

Shazam has 70 million active users every month and in just one year (2012) it generated $300 million in digital sales, primarily through iTunes. It also has approximately 35 million of songs in its database [4].

2 Content-based Music Information Retrieval

One of the Shazam's founders, Avery Li-Chun Wang, describes the purpose and the idea behind this highly popular application in his PhD thesis [5]. He was interested in how we, as psychological as well as biological organisms, perceive sounds, and how an automatic system for music identification could be developed.

Such a system would take a recorded sound and automatically separate it into representations of component auditory streams which could later be played back in isolation.

Wang's idea goes along the same lines as the Kurzweil's [6] statement that "Like a lot of clever computer software, it was a solution in search of a problem". Becoming part of the founding team of Shazam, Wang was able to develop a mobile music identification system, which in that time was still in search for a problem to solve. In his research paper on industrial-strength audio search algorithm [7], he explained procedures needed to recognize music on the smartphone with Internet connection.

The aim of the research presented in this paper was to evaluate how precise Shazam is in two different environments, the first one being a night club (where the music is loud and the signal to noise ratio is high) and the other one being home studio environment, where the signal to noise ratio is low.

Furthermore, commercial use of the patents covering Shazam algorithms will be discussed.

In the year 2005, Broadcast Music, Inc. (BMI) acquired the technology and patents from Shazam Entertainment, creating Landmark Digital Services [8]. BMI wanted to apply the same technology to monitor public performances for the purpose of more accurate distribution of royalties to music publishers.

According to Casey et al [9], there are different strategies to approach this relatively new scientific field. As more and more music is created, there is a constant need to develop technology which will help us manage such large music collections, which are also distributed to a number of markets in various media formats, from vinyl to digital audio streaming on mobile phones.

There are two major possible ways to manage such a big amount of content: one is by using meta-data, while the other one implies content based search [10]. The problem with the metadata is that its creation could be extremely time consuming and expensive and, according to Casey [9], to enter metadata for one million of songs could take approximately 50 person-years.

According to Casey [9], content based MIR could be divided to:

- high-specificity systems that match instances of audio signal content;
- mid-specificity systems that match high-level music features, such as melody, but do not match audio content and
- low-specificity systems that match global (statistical) properties of the query

In the context of Music Information Retrieval, the concept of specificity denotes how discriminating a particular task is, or how clear is the dividing line between relevant and irrelevant retrieval results [11, 12].

In other words, specificity determines the amount of acoustic and musical material in a retrieved result that needs to be shared with a query for a result to be considered relevant and the number of documents in total that could be considered relevant retrieval results.

Music recognition, which is the main topic of this paper, is defined as a highest possible specificity. Shazam is the application which fulfils the above mentioned music recognition tasks.

3 Transaction cost Theory and Music Industry

If the whole is more than sum of its parts, then combining two theories or scientific fields could provide us with a basis for a system which could benefit everybody involved. Therefore, in this paper we would like to address two theories: the Transaction Cost theory and the theory of Music Information Retrieval in the context of music industry, specifically in the context of income from music rights (royalties).

In their strategic guide to the network economy [13] Shapiro and Varian state: "As an owner of an information good, you should ask yourself: is it cheaper for me to distribute my products directly to the end user or is it cheaper for the organization to distribute my product to the end user".

Authors further propose that if the organization and owner share gains of effectiveness of such a system, they are able to reach bargain. Such a gains and advantages are based on the Transaction Cost theory, probably one of the most significant theories related to the information age.

Already Coase [14] concluded that the cost of the transaction regulates the size of companies: if a cost to outsource a service is cheaper, then there is no reason to do the transaction inside the company.

Accounting services for successful small companies make a good example principle: it is cheaper for them to have such services outsourced, then to employ accountant in the company.

Speaking precisely in terms of collecting music royalties, it would be expensive if every copyright owner tries to issue an invoice for use of its copyrighted property to every party using it [10]. This is the basic economic reason why music authors' collecting societies emerged in the last century and are active on the music market.

"An audit revealed more than 20 separate areas where Capitol/EMI had "wrongfully accounted" for costs or revenue concerning promotion, manufacture and sales, resulting in $19 million of unpaid royalties due the Beatles from 1969-1979" [15].

As we could see from the above discussed examples, the internal transaction cost can play a significant role in music information retrieval research.

But, as Ronald Coase proposed in his lecture [16] citing Thomas Kuhn's book "The Structure of Scientific Revolutions", scientists aren't receptive to proposals for a change in their subject unless they are dissatisfied with the old views.

Reflecting on that statement, we recognize an emerging need to make information sciences (i.e. music information retrieval) and economy working closely together, as both fields deal with a system and ways information is communicated through that system.

We could assume that information technology was not well developed in 60's and 70's of the 20th century, but in the contemporary environment, we should ask ourselves if technology can actually help businesses to become better organised by providing infrastructure for the information flow.

If we take a closer look to the income of the top-ranked rock performer in 2002 [17], Paul McCartney, we can discover that he received $64.9 million from live concerts, $2.2 million from recording sales and $2.2 million from copyright royalties. The figures reveal that he received the same amount of money from copyright royalties as from record sales.

Furthermore, according to the ratio extracted from International Survey of Music Publishing Revenues in 2001 (published in Connolly and Krueger's paper [15], 23.6 % of the revenues from music publishing in the U.S. came from live performance. In the same report [10], the sum of the publishing income in top 10 countries of the world reached 2,619.05 million of US dollars in 2001 only.

Based on the proposed ratio, if we calculate the amount of the revenues that were collected from live performances (e.g. concert venues or other public spaces), we come up to 618.09 million of US dollars.

In the research conducted for the current study, authors could not find any empirical evidence or articles on how the process of collection and distribution is executed, but Kretschmer [18] provides a great insight into distribution among copyright owners. Findings related to our topic show that in UK in 1994 only 1.96% of copyright owners made more than 20,000 GBP annually, while in the same time in Germany 3% of copyright owners made more than 25,000 GBP annually.

Our research never aimed to offer a detail economic analyses or guidelines for the development of the methodology for such analysis. Our goal was to question MIR technology and its potential to improve effectiveness of an organisation.

The results presented here are only a small scale preliminary findings, trying to answer the following question: could the technology that is available today (more precisely: applications such as Shazam) improve organisational systems of performing rights organizations (PRO) in the area of precise monitoring of the songs used in the night club environment?

In other words, the aim of our research was to prove whether the technology can help to reduce transaction costs and improve efficiency and transparency of PRO’s.

A discussion on the details of different perspectives regarding economical and information technology issues related to this research will be given in the discussion chapter.

4 Experiment

The small scale research described in this chapter provided us with preliminary results which could be used as a base for the development of more extensive research in this field. The idea behind our experiment was quite simple: we aimed to check how precise Shazam application is in recognising the songs played in nightclubs.

Since electronic music became a popular force in music markets, big DJ stars earn today as much as rock stars earned back in the days [19]. However, scientific evidence on how playlists in nightclubs (and other public spaces) are collected, administer and distributed does not exist (at least not publicly). Therefore, we decided to test Shazam in the night club.

In our experiment, each song was tagged for 10 seconds in two different environments and Shazam created two different audio fingerprints for each song based on some of the anchors of the simplified spectrogram and the target area between them.

For each point of the target area, Shazam created a hash value that is the combination of the frequency at which the anchor point is located, the frequency at which the point in the target zone is located, and the time difference between the point in the target zone and when the anchor point is located in the song.

While we were trying to perform the experiment in the night club environment (the first experiment), one examiner was standing next to DJ, taking notes of the songs he played, while the other examiner was on the dance floor holding two different smart phones: Galaxy1, and iPhone5 (both having Shazam installed) clicking the Shazam “listening button“ and recording tags in "My Tags" list. Galaxy1 used Blue version of Shazam application, while the version installed on the iPhone was Red Version.

After the first experiment, the playlist was transferred to the spreadsheet and links to music were found on the internet for the comparison in the home studio environment. The same approach was used for the second experiment in the home studio environment.

Examiners have also taken notes on all different songs that Shazam retrieved as a result of a query. It’s important to point out that DJ in the club was asked to play music that is already available on the market and not to play new music, demos or promotional copies of songs which are not yet published.

Sound system that was used in the club was: Electro Voice 2 x MTL2 Subs, 4 x Funktion one res 4 flying from truss, 2 x Funktion F18 monitors.

The experiment was performed on a 5 meter distance from the sound system. In home studio consisted of MacBook (MacBook 7.1, Intel Core 2 Duo, 2.4 GHz) and songs were analysed from speakers which are installed on that computer.

Percentage of songs recognised by Galaxy in the club environment	5.00%
Percentage of songs recognised by iPhone in the club environment	45.00%

Table 1. Shazam music recognition results for iPhone5 and Galaxy1 in the public place

Percentage of songs recognised by Galaxy in the home studio environment	70.00%
Percentage of songs recognised by iPhone in the home studio environment	75.00%

Table 2. Shazam music recognition results for iPhone5 and Galaxy1 in the home studio

Galaxy in Home Studio Environment (accuracy)	25%
iPhone in Home Studio Environment (accuracy)	55%

Table 3. The accuracy of song recognition

Galaxy in Home Studio Environment (14 songs in total recognised)	18:22:00
iPhone in Home Studio Environment (15 songs in total recognised)	7:43:00

Table 4. The total time Shazam needed to accurately recognise all songs

The above listed tables (Table 1 and Table 2) present the result of our experiment, showing the percentage of songs from the same playlist recognized in two different environments (night

ISBN: 978-960-474-361-2

club and home studio) by two different mobile devices (smartphone Galaxy1 and iPhone5).

Furthermore, although Shazam managed to recognize more than 70% of songs in the home studio environment, for some queries it firstly offered one or more songs that weren't on the playlist as a result, before recognizing the right one. Table 3 reveals the percentage of songs that were accurately recognized in the first run.

Finally, a detailed playlist and the summary of the research results of our experiment are given in the Table 5.

In the following table (Table 5) we provide the information on Artist, Song and whether the song was recognized by Galaxy smartphone (G) or iPhone (I) in the night club or at the home studio environment (0= not recognized; 1 = recognized).

Finally, Table 5 shows the total time Shazam needed to recognize the song on Galaxy smartphone (G) or iPhone (I), for songs the application managed to recognize.

		Night club		Home		Recognition time	
Artists	Track	G	I	G	I	G	I
Fresh and Low	No Going Back	0	0	1	1	1:30	1:20
Ivano Tetelepta	Smokin G	0	0	1	1	4:24	1:41
Kelvin K	Ancestral Moon	0	0	1	1	0:24	0:18
Ark	Nuark	0	0	0	0		
Terrry Lee Brown Jr	Our Rhythm	0	0	0	0		
Highland Brothers Inc.	This One's A Keeper	1	1	1	1	0:48	0:14
MP	The Domm	0	0	0	0		
Timewriter	Love Trap	0	1	1	1	3:37	0:13
Kelvin K	G'groove	0	1	1	1	1:13	0:34
Basement Jaxx	Fly Life	0	1	1	1	1:29	0:12
Point G	Braka	0	0	0	0		
J.C.	Differential	0	0	0	0		
Cevin Fisher	The Way We Used To	0	1	1	1	0:35	0:28
E.B.E.	Werked	0	1	1	1	1:43	0:27
Van Basten	Blood wars	0	0	0	0		
Sasse	Soulsounds	0	1	1	1	0:44	0:30
DJ Vasile	Nu vrei sa mergi?	0	0	0	0		
Atom	Love To Heart (Too Hot)	0	0	1	1	0:20	0:24
Scoper And Bubba	I'm Satisfied	0	1	1	1	2:39	0:15
Birdsmakingmachine	Black Pearl	0	1	1	1	1:12	0:19
Caucasian Boy	Northern Lighst	0	0	1	1	1:04	0:24

Table 5.[1] The playlist and the results for each song

5 Discussion

Given above presented numbers, we can conclude that Shazam application is not very accurate in the night club environment and, at this stage, cannot be used by performing rights organizations to monitor the music played in public spaces.

We also have to emphasize that the accuracy of song recognition was very low, but this could be explained with the fact that Shazam analysed songs from the very beginning and dance music compositions have long intro parts, so they could be mixed more smoothly by DJ.

Furthermore, we would like to discuss some issues that emerged during the research.

The role of PRO's in the overall system of the music industry is of high importance, since it's economically ineffective for a copyright holder to trace how his/her intellectual property is being used.

But, as income from copyright tends to be almost the same as the income from music sales, it would be interesting to compare the amount of the money invested in the research and development of media technologies that carry music with the amount of the money invested in technology used to support monitoring, collecting and distribution of the income from copyrights.

If PRO's provide the income for only 1.96% of UK artists and 3% of German artists, it would be fair to thoroughly evaluate the economic model of such a system.

[1] I = iPhone5, G= Galaxy1, 1=recognised, 0=not recognised

ISBN: 978-960-474-361-2

Although Shazam is a successful commercial company, our preliminary research revealed that their algorithm is still not efficient enough to monitor the use of the copyrighted material in public spaces.

In order for royalties to be collected and distributed correctly, one needs to have precise instruments, that are also cheap and transparent in terms of information processing and that offer the guarantee that music creators will get their due.

6 Conclusion

Apart for benefit of such applications for the end music users, applications such as Shazam could be used to identify the music being played in public spaces, which are required to pay music authors' collecting societies for the music distribution.

Therefore, our conclusion is that at least some research funding should be devoted to the development of new technologies for monitoring music played in public spaces. Also, such research activities should be conducted involving all stakeholders: artists, PRO's and public space owners.

We conclude this paper by indicating the area of applicability for music identification applications. It is a known fact that the first PRO in the U.S. was ASCAP[2], which was established in 1914, but the very first money the artists received from royalties was in 1921. Today, we live in a society that can easily develop technologies to accurately monitor music played in an economical and transparent way, to the benefit of all music creators, music publishers, licensees, collecting societies and music consumers.

Acknowledgements:
We would like to thank DJ Valentino Kanzyani for the support in this research and Nick Colgan from Barbarellas Discotheque, Tisno, Croatia.

References:

[1] Woodward, D. (2000, 11. December). Shazam names that tune, *Director*. Retrieved from http://www.director.co.uk/magazine/2009/11%20December/shazam_63_04.html

[2] Mance, H. (2013, 29. April), App producer Shazam aims for $1bn IPO, *Financial Times*. Retrieved from http://www.ft.com/intl/cms/s/0/d07809fa-b0f0-11e2-9f24-0144feabdc0.html#axzz2naJQhdFN

[3] Shazam (service). (n.d.) in *Wikipedia*. Retrieved September 13, 2013, from http://en.wikipedia.org/wiki/Shazam_(service)

[4] Dotted Music (2013, 19. September), SAtN11: Being Discovered With Shazam (Jon Davies, Partnerships Manager) [Video file]. Retrieved from http://www.youtube.com/watch?v=VopGy3pO_wI

[5] Wang, L.A. (1994), Instantaneous and Frequency-Warped Signal Processing Techniques for Auditory Source Separation. Ph.D Thesis, Stanford University.

[6] Kurzweil, R. (2001), The Age of Spiritual Machines: When Computers Exceed Human Intelligence. *Penguin* (Non-Classics), First edition.

[7] Wang, L.A. (2003), An industrial-strength audio search algorithm. In: Proceedings of ISMIR 2003, 4th Symposium Conference on Music Information Retrieval, 2003. pp. 7-13. Available: http://www.ee.columbia.edu/~dpwe/papers/Wang03-shazam.pdf

[8] Inman, D. (2010, 13. September), Next BIG Nashville Spotlight: Landmark Digital Services, *American Songwriter*. Retrieved from http://www.americansongwriter.com/2010/09/next-big-nashville-spotlight-landmark-digital-services/

[9] Casey, M.A.; Veltkamp, R.; Goto, M.; Leman, M.; Rhodes, C.; Slaney, M (2008). Content-Based Music Information Retrieval: Current Directions and Future Challenges. Proceedings of the IEEE. pp. 668 – 696.

[10] Rhodes, C.S., Rauchas, S. (2009) Creative computing II: interactive multimedia. Volume 2: Perception, multimedia information retrieval and animation, pp. 51-61. Available: http://doc.gold.ac.uk/~mas01cr/teaching/is52020b/2009-10/cc227-vol2.pdf

[11] Müller, M. (2007) Information Retrieval for Music and Motion. Berlin: Springer-Verlag. ISBN 9783540740476.

[12] Witten, I.H., Moffat, A., Bell, T.C. (1999) Managing Gigabytes: Compressing and Indexing Documents and Images. San Francisco: Morgan Kaufmann Publishing. ISBN 1558605703.

[13] Shapiro, C., Varian, H. (1998). Information Rules: A Strategic Guide to the Network Economy. Cambridge, MA: Harvard Business School Press.

[14] Coase, Ronald (1937). The Nature of the Firm. Economica (Blackwell Publishing) 4(16). pp. 386–405.

[15] Connolly, M., Krueger, A.B. (2007) Rockonomics: The Economics of Popular Music. The Milken Institute Review 9(3). pp. 50–66.

[2] American Society of Composers, Authors and Publishers

[16] Coase, R. (2002, 4. April), Why Economics Will Change, Remarks at the *University of Missouri, Columbia,* Missouri, USA. Retrieved from http://www.coase.org/coaseremarks2002.htm

[17] La Franco, R. (2003, April), The 2nd Annual Richest Rock star list, *Rolling Stone* magazine, #919, n. pag.

[18] Kretschmer, M. (2005), Artists' earnings and copyright: A review of British and German music industry data in the context of digital technologies. First Monday 10 (1). Available at: http://firstmonday.org/ojs/index.php/fm/article/view/1200/1120

[19] O'Malley Greenburg, Z. (2013, 14. August), Electronic Cash Kings 2013: The World's Highest-Paid DJs, Forbes. Retrieved from http://www.forbes.com/sites/zackomalleygreenburg/2013/08/14/electronic-cash-kings-2013-the-worlds-highest-paid-djs/