

Speaker Recognition Will be Running Everywhere

Duowei Tang, Miaomiao Wu

When it comes to biometrics, the very first thing that comes to us may be fingerprints. However, we do actually have our own unique voiceprint as well, which is similar meaning like fingerprints. It is the unique part in people's voice instead of fingers. Voiceprint recognition, which is also called speaker recognition, has gain more and more attention nowadays. In our eyes, speaker recognition technology will gradually gain its appealing and will become more and more widely used in the future. We believe that the speaker recognition technologies will be running in many kinds of domains and it will influence our life a lot or even change it in around 5 to 10 years.

1. Introduction

We usually unlock a system or log in our account with a password or fingerprint and face scanning in some high-tech situations. Speaker recognition, is another possibility. It recognize a person by voice or voiceprint more precisely. During several decades of researching, scientists find some features from one's voice can be seen to uniquely represent a person. In training process, a piece of voice is processed to extract the features, then base on those features, we can use mathematical ways to model a person. Finally in testing procedure, testing voice is also extracted to get its features, after comparing those features with training features, the most possible answer can be got through statistical calculations.

2. Advantages and Applications

There are many benefits of using speaker recognition. Comparing to fingerprint recognition, the data of speaker recognition is collected through a fingerprint scanner which is more expensive and more complex than a microphone which is enough for speaker recognition. On the other hand, voice can be acquired remotely, so speaker recognition is not restricted by distance. For example, it provide the possibilities that recognize a person through smartphone or fix lines.

An attracted scenario was announced by Wang Qingwen, the inventor of Roslink smart home system where speaker recognition is used in the door lock (Xia, 2014). It will provide much more convince when a lady back home after shopping with her hands full of goods, she can easily open the door be speak a "Password" instead of looking for her key and make everything a mess. Another example of its application is about people living condition investment in pension insurance in China. Since China have the enormous amount of people that almost all families have 4-2-1 structure that is one child, two parents and 4 grandparents in another word. The traditional way of doing that is visit every families and find out the living conditions of the people, which is a huge work and the working efficiency is very low. Fingerprint collection can

be used as a survival certificate, however, it is difficult to implement since it needs a certain device in every families to collect the data. In this situation, speaker recognition technique can give a better solution to solve this problem. The voice can be collected simply by phone. It doesn't need any extra settings or any form of interview which is done in a fixed office.

Speaker recognition also gives another aspect to describe a people, which can be used for example in criminal investigation. A good proof was in February of 2015, a new reported a phone fraud where a man mimic a women's friend voice, and fraud for 15 thousand RMB. When investigation, speaker recognition played a very important role. At beginning, there was not enough proof that ensures the criminal suspects, then the police got the phone voice record and use speaker recognition techniques and finally arrested the criminal. (Cao, 2015)

Speaker recognition is easy to use, easy to be integrated. Since people are less and less rely on complex interact with devices, voice control seems a very important trend that the smart device will be developed. Speaker recognition is very easy to be integrate with voice control, it doesn't need multiple acquisition ending devices, and the voice front end processing is quite common in both speaker recognition and voice control which simplify design a lot. It can be applied directly in existing fix line and mobile phone system, we don't need install any new devices.

3. Development in Recent Years and Prediction

With those benefits and huge potential values, people are also paying more and more attention on researching and developing about it. There is a report about the number of patents recorded in last few years and the prediction in following years in China (Figure 1). The blue line is the prediction, and the green bars are the records before 2013. We can easily see we are now in the knee point of the S shape curve, and in the next following years, speaker recognition will increase much faster although it might saturate in some point. However, it gives us a direct view that how important the speaker recognition is and what a good vision can we expect from it.

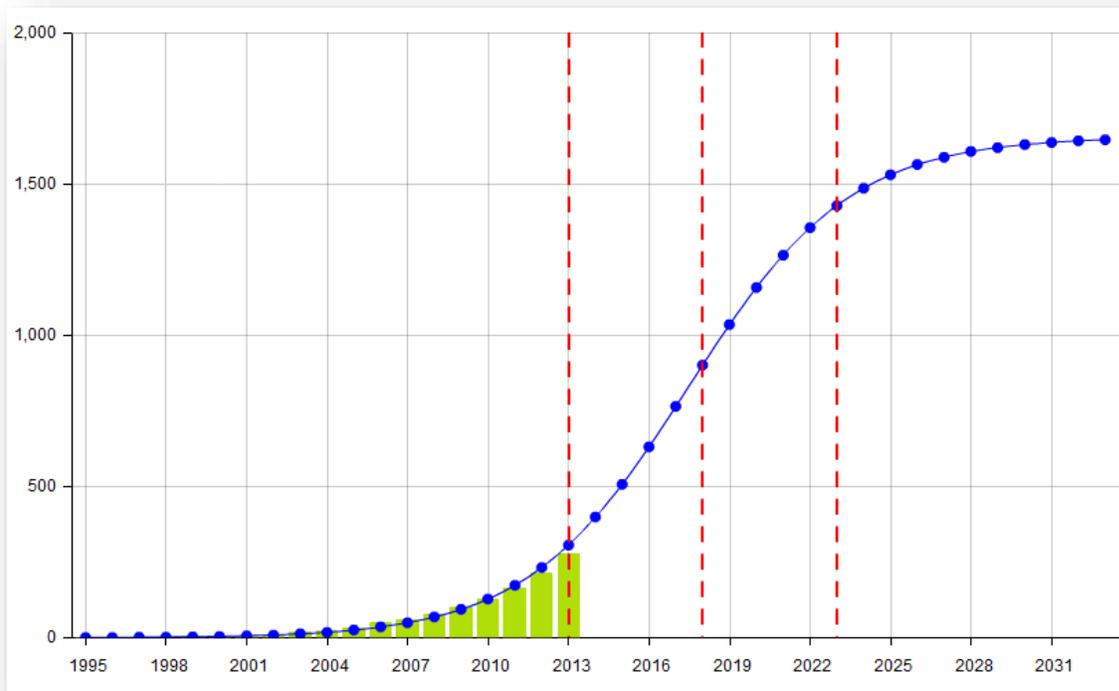


Figure 1 the prediction of the number of Speaker recognition patents (TEKGLORY, 2013)

On the other hand, Industry 4.0 is proposed recently, which characteristic for industrial production in an Industry 4.0 environment are the strong customization of products under the conditions of high flexibilized (mass-) production. (Wikipedia, Industry 4.0, 2015) In this general idea, there is a key point that is Decentralization where smart factories or machines will make decision on their own. It is not difficult to find the opportunity that speaker recognition will play a very important role here and it provide a wider possibility to make a machine more “Human Like”. Additionally, it could be the basic block of a whole smart system. We can also find that it could contribute a lot for the whole decision making process. In this background, speaker recognition must be treated as a VIP, and it won’t miss the chance that running on every Industry 4.0 machines.

Companies also afraid of losing the good opportunity to have a position in speaker recognition field. Nuance Communications is an American multinational computer software technology corporation who introduced "Nuance FreeSpeech" verification system to Barclay phone verification system. Now the Barclay phone verification system can use biometric technology to authenticate via voice. The voice biometrics will obtain data from a 30s conversation between call center employees and customers, and the accuracy is over 95%. They declare that over 84% of customers are trying this new service. This technology is expected to quickly cover the United States, so they do not need to call SSNs when every time they call service center. Furthermore, in 2014, ING introduced a voice-controlled mobile banking app. This app was also powered by Nuance (Wikipedia, Nuance Communications, 2015).

Those phenomenon reveals the vital position of speaker recognition's in people's mind, which also make us have more confidence in it. We believe that speaker recognition will have a good future and we will see it runs everywhere, and change the world. Besides, some applications of it have already came up. The best example is its application on WeChat, one of the most popular social Apps in China. Last year, WeChat released its "Voice Lock". "Voice Lock" can be a choice to login your account instead of type a long password.

In our blog, we mentioned the intelligence will grow up in a logarithm curve for the simple reason that faster machines increase our working efficiency very much, it means that we can already do something that make the world more intelligent in an efficient way (Urban, 2015) . For example, if time backs to 100 years ago, we can never knew the fact that in next decades there will be computers everywhere in this planet. However, only in 5 years, from 2010, almost everyone holds a smart phone! Everyday new inventions come up, and machine gradually have their "Brain", we cannot imagine, how fast the intelligence will increases as if all the machine can also "Thinking". This truth also show in another side how fast will speaker recognition technologies being developed, which is an essential part of artificial intelligent.

4. Limitations and Challenges

But everything have both sides. Speaker recognition do give us conveniences, in the other hand, it still have its own limitations and threaten. In our blog, most of people doubting that what if someone record your voice and use it for the recognition, will it still be safe? We can also propose some other situation this technique have to face to, like what if someone mimic your voice, or someone use your voice model to generate very similar machine voice? Those are the possibilities that will make the system failed. Luckily, the solution is already there. Since when you use different device to play a voice or you mimic a voice, the channel will be different than the original one, some technique which used to detect channel difference will be useful here.

Accuracy is also a bottleneck that limit this technology. Since going for high accuracy means more data is needed for modeling and testing, which will cost more calculation time. In addition, noise will also have big influence on the performance. It seems that a noise reduction technique and high speed computer is urgently required here. The recently research on this topic nowadays indeed focus on how to reduce the dependency of the number of training and testing data, like the state-of-art technique i-Vector analysis which is also involved in our thesis research. In short, a single speaker recognition system may not work properly at all conditions, it needs the help of all technical aspects. We have to admit that there is still a long way to go.

Information safety is more important for people's privacy. When using speaker recognition, a huge database is needed to set up by recording the users' voices. Those voices contains people's unique information and it will be very dangerous if those information is leaked. In the future times when voice lock is used everywhere, in banks, in rooms and in accounts logins, voiceprint will be the highly secret information. To avoid this, data encryption may works. For example, encrypt the model data that stored in memory. Encryption technologies are still a temporary solution, it needs updated along with the fast increasing of calculation speed. However, we should protecting our information security by knowing how it will be leaked, and rise our level

about understanding of that and action against it. Even more important is that we also need everybody be aware of educating people don't do criminal things from that. In our opinion, if all the people in the world use those technologies only in a positive way, it will be developed even faster, and the benefits will become more for us as a human being.

5. Promising Future

It is a great pleasure to see so many great companies and teams are devoting themselves to the research and development of speaker recognition technologies. In conclusion, speaker recognition will be a hot topic in following years, we can see its very good benefits and we can feel the convenience it give us. "My voice is my password", the slogan shows that how easy and elegant it will make our life be. Although there still remains some barrier that lower the developing speed of speaker recognition, we are expecting that speaker recognition will change the world from touching to voice control, and speaker recognition itself will witness a next information revolution. We firmly believe that it will become a very vital part of our life and it will run in every smart devices.

References

- Cao, L. (2015, 2 3). *Fraud money by mimic women's friend's voice, speaker recognition find the criminal*. Retrieved from E23.cn: <http://news.e23.cn/content/2015-02-03/2015020300386.html>
- TEKGLORY. (2013). *TEKGLORY Patent Trend Prediction Report (# 40)*. Retrieved from TEKGLORY: <http://www.tek-glory.com/about/news/20140701/138.html>
- Urban, T. (2015, 1). *The AI Revolution: The Road to Superintelligence*. Retrieved from WAIT BUT WHY: <http://waitbutwhy.com/2015/01/artificial-intelligence-revolution-1.html>
- Wikipedia. (2015, 3 29). *Industry 4.0*. Retrieved from Wikipedia: http://en.wikipedia.org/wiki/Industry_4.0
- Wikipedia. (2015, 4 11). *Nuance Communications*. Retrieved from Wikipedia: http://en.wikipedia.org/wiki/Nuance_Communications
- Xia, H. (2014, 10 10). *Roslink smart home system: combined with speech and speaker recogniton, make home smarter*. Retrieved from CYZONE: <http://www.cyzone.cn/a/20141010/264113.html>