

The online speech recognition feature of the ABSONIC speech recording, speech management and speech processing system using the speech-to-text subscriptions of GOOGLE and/or NUANCE covering more than 100 languages and dialects (tariff of GOOGLE – 10.07.2017: approx. 0.024 USD/min. ~ 1.44 USD/hour ~ 1440 USD/1000 hours)

1. Information on the basic ABSONIC system without speech recognition available at:
http://www.absonic.co.uk/download/technical_factsheet.pdf
2. An online speech recognition feature has been added to the ABSONIC speech recording, speech management and speech processing system, allowing for live speech (dictation, meetings, lectures, language learning/teaching, infocommunication exchanges) to be automatically displayed and/or transcribed in real time or also later on by using stored speech records. This integrated package comes as actually a boxed product without the special features of ABSONIC that would otherwise require other devices, accessories and/or support:
http://www.absonic.co.uk/index.php?x= EN/factsheet_en
3. Speech recognition by GOOGLE and NUANCE is done in remote, unknown countries so the information security of the speech or text material sent is not at the same level as with speech recognition on your own computer or server. **For this, upon request and after the necessary preparation, we provide a solution in Hungarian perhaps English with our own ABSONIC speech recognition software (It utilises probably weeks or months).**
4. The accuracy of speech recognition depends on the speech recognition accuracy of GOOGLE and/or NUANCE. So ABSONIC only provides the means for speech recognition by providing the control tools and the connection but does not influence the accuracy or interferes with the process of speech recognition. For this reason, the ABSONIC system has no responsibility at all in the following:
 - 4.1. any issues related to service providers
 - 4.2. the accuracy of speech recognition
 - 4.3. issues of registration
 - 4.4. the quality of the service
 - 4.5. payment issues
 - 4.6. changes in the service provided (which we intend to follow up on and maintain through our central updating system if the service providers make it possible)
 - 4.7. the system of GOOGLE is just a beta version currently, so it is not final and there still might be significant changes, although it is already practical and easy to use.
5. The speech recognition accuracy of GOOGLE and/or NUANCE depends on the below factors, without wanting or being able to provide an exhaustive list (part of the below issues are compensated for to a degree in the following: noise filtering, automatic gain control, echo cancellation etc.):
 - 5.1. speaking abilities of the speaker (volume, voice quality, errors of speech, etc.)
 - 5.2. the speaker's knowledge of the language
 - 5.3. the acoustic conditions and noise levels of the speaker's environment
 - 5.4. the quality of the speech sensor device: microphone and/or infocommunication speech set
 - 5.5. the distance between the speaker and the speech recognition device (microphone, infocommunication speech set)
 - 5.6. the quality of the speech forwarding system of the speech sensor device (speech transmission devices, channels etc.): microphone networks, infocommunication networks, data forwarding and processing units, etc.
 - 5.7. the quality of the internet connection used for speech recognition
 - 5.8. the level of sophistication and accuracy of speech recognition more than 100 languages and dialects covered by GOOGLE and NUANCE might vary. From the several languages of the two service provider, one part of languages/dialects are overlap, meaning they can be found at both providers. The remaining languages/dialects are available with one of the providers only.
 - 5.9. Upon request, we can significantly reduce, influence some of the above listed, location specific negative factors following preliminary coordination and on-site inspection as part of a separate agreement.

6. Live speech can also be recognized real time with some delay (in a rough estimate: 2-10 seconds on average, maybe more), which cannot be precisely determined in advance (it requires location specific user tests):
 - 6.1. instant display and transcription of dictations
 - 6.2. instant display and transcription of meetings
 - 6.3. real time display of lectures on screens in lecture halls and on the computers of the audience
 - 6.4. language teaching and language learning
 - 6.5. instant display and transcription of infocommunication exchanges
7. In case of the transcription of stored speech records, there is no need for real time display as with live speech so the time required for the transcription of such records essentially depends on the length of the stored speech record itself.
8. The recognized text can be edited in RTF format using the text editor built into ABSONIC.
9. The speech record and the recognized text can be managed together or separately (editing, exporting-importing, archiving etc.).
10. The text created as a result of speech recognition can be transferred to the fee paying, online translation software of GOOGLE as soon as we have finished this development module.
11. The ABSONIC systems, currently, allows for running one speech recognition at a time, although it is a multichannel system. The development of simultaneous speech recognition in multiple languages on multiple channels is currently under development. This might mean several microphone channels (e.g. for interpreters) or tens, even hundreds of infocommunication channels: analogue phones, ISDN, VoIP, GSM, Viber, WhatsApp, Messenger, satellite phones- INMARSAT-IRIDIUM-THURAYA, etc.
12. The regularly updated language and dialect set of GOOGLE language can be found at the following link (current ones attached: 10.07.2017.), based on this you can test and determine which service provider you wish to use and for what jobs:

<https://cloud.google.com/speech/docs/languages>

 - 12.1. For subscription and registration, a GOOGLE account is required.
 - 12.2. For managing contact with the service provider including registration, sufficient knowledge of English and computing skills (web related) are required. All activities, transactions, events resulting from connecting to the service provider are the sole and full responsibility of the registered user in every respect.
 - 12.3. Subscription is subject to registration requiring the provision of bank card data:

<https://console.cloud.google.com/> - in the upper left corner: Sign up for free trial – Giving the country, then the answer is ‘yes’ to both questions. – On the next page the name of the company and/or personal, bank card details need to be provided.
 - 12.4. Payment is done according to the time used on a periodical (monthly) basis by debiting the bank card after having used up free credits.
 - 12.5. 0-60 minutes of speech recognition per month is free.
 - 12.6. From the 60th minute onwards, the fee is 0.006 USD/15 seconds = 0.024 USD/minute ~ 1,44 USD/hour ~ 1440 USD/1000 hours.
 - 12.7. For the first time, for the first year, all registered users receive 300 USD ~12500 minutes ~ 200 hours of free credits that can be used for a year in the first year.
 - 12.8. The tariff of GOOGLE can change, so you have to controll regularly.
13. The regularly updated language and dialect set of NUANCE can be found at the following link (current ones attached: 10.07.2017.), based on this you can test and determine which service provider you wish to use and for what jobs:

<https://developer.nuance.com/public/index.php?task=supportedLanguages>

 - 13.1. Subscription is subject to registration:

<https://developer.nuance.com/public/index.php?task=prodStart>
 - 13.2. For managing contact with the service provider including registration, sufficient knowledge of English and computing skills (web related) are required. All activities, transactions, events resulting from connecting to the are the sole and full responsibility of the registered user in every respect
 - 13.3. Prepayment is done via PayPal and the connected bank card or bank account (e-mail address and password). The minimum amount of prepayment is 25 USD equalling 3125

transactions or sentences. Later payments can be done on an as needed basis personally or via direct debit, potentially varying from country to country.

13.4. The recognition of 20000 transactions, i.e. sentences, per month is free. Above this level the fee is 0.008 USD/transaction or sentence. The system breaks down the text into sentences according to the pauses in speech. Based on a rough, statistical estimation, the maximum length of a sentence is 30 seconds because this is the time duration for which a person can talk continuously without taking breath. But the actual average duration of a sentence is 10 seconds. Using these approximations, 20.000 transactions/sentences equal about 10.000-30.000 minutes, 167-500 hours of speech free of charge per month. Note: If speech is so fast that the pauses between the sentences are not detectable properly or if for any other reason the sentences cannot be clearly defined, the transactions are determined according to their length where the maximum duration of a transaction 2 Mbyte equalling about 1 (one) minute of 16-bit, 16-kHz PCM mono audio signal or speech. In this case speech recognition and its costing is done in such data packages.

13.5. It is possible to upload further entries into the dictionary for words not known by the speech recognition system using files (WORD, TXT, etc.) :

<https://developer.nuance.com/public/index.php?task=uploadVocabulary>

13.6. The tariff of NUANCE can change, so you have to controll regularly.

14. Should you have any questions, recommendations or opinion on the above, please do not hesitate to contact us at:

Dr. Antal Miklos Varhalmi infocommunication engineer-expert

email1: miklos.varhalmi@nicopro.us

email2: miklos.varhalmi@absonic.co.uk

email3: varhalmi.miklos@nicopro.hu

homepage1: www.absonic.us

homepage2: www.absonic.co.uk

homepage3: www.absonic.eu

GOOGLE and NUANCE languages-2017-07-10				
No	Country	Languages	Google	Nuance
1.	Algeria	Arabic	x	x
2.	Argentina	Spanish	x	x
3.	Australia	English	x	x
4.	Austria	German		x
5.	Bahrain	Arabic	x	x
6.	Bangladesh	Bengal		x
7.	Belgium	Dutch		x
8.	Belgium	French		x
9.	Bolivia	Spanish	x	
10.	Brazil	Portuguese	x	x
11.	Bulgaria	Bulgarian	x	x
12.	Canada	English	x	x
13.	Canada	French	x	x
14.	Canada	Hindi		x
15.	Chile	Spanish	x	
16.	China	Chinese, Mandarin (Simplified)	x	x
17.	Colombia	Spanish	x	x
18.	Costa Rica	Spanish	x	
19.	Croatia	Croatian	x	x
20.	Czech Republic	Czech	x	x
21.	Denmark	Danish	x	x
22.	Dominican Republic	Spanish	x	
23.	Ecuador	Spanish	x	
24.	Egypt	Arabic	x	x
25.	El Salvador	Spanish	x	
26.	Finland	Finnish	x	x
27.	France	French	x	x
28.	Germany	German	x	x
29.	Greece	Greek	x	x
30.	Guatemala	Spanish	x	
31.	Honduras	Spanish	x	
32.	Hong Kong	Chinese, Cantonese (Traditional)		x
33.	Hong Kong	Chinese, Mandarin (Simplified)	x	x
34.	Hungary	Hungarian	x	x
35.	Iceland	Icelandic	x	
36.	India	English	x	x
37.	India	Assamese		x
38.	India	Bhojpuri		x
39.	India	Bengali		x
40.	India	Gujarati		x
41.	India	Hindi	x	x
42.	India	Marathy		x
43.	India	Oryja		x
44.	India	Punjabi		x
45.	India	Telugu		x

No	Country	Languages	Google	Nuance
46.	India	Urdu		x
47.	India, Malayzia	Tamil		x
48.	Indonesia	Bahasa	x	x
49.	Indonesia	Indonesian		
50.	International	Arabic		x
51.	Iran	Persian		x
52.	Iraq	Arabic	x	x
53.	Ireland	English	x	
54.	Israel	Arabic	x	
55.	Israel	Hebrew	x	x
56.	Italy	Italian	x	x
57.	Japan	Japanese	x	x
58.	Jordan	Arabic	x	
59.	Kuwait	Arabic	x	x
60.	Latin-America	Spanish		x
61.	Lebanon	Arabic	x	x
62.	Lithuania	Lithuanian	x	
63.	Malaysia	Malay	x	x
64.	Malayzia	Hindi		x
65.	Mexico	Spanish	x	x
66.	Morocco	Arabic	x	x
67.	Nepal	Nepali		
68.	Netherlands	Dutch	x	x
69.	New Zealand	English	x	
70.	Nicaragua	Spanish	x	
71.	Norway	Norwegian Bokmål	x	x
72.	Oman	Arabic	x	x
73.	Pakistan	Urdu		
74.	Panama	Spanish	x	
75.	Paraguay	Spanish	x	
76.	Peru	Spanish	x	x
77.	Philippines	English	x	
78.	Philippines	Filipino	x	
79.	Poland	Polish	x	x
80.	Portugal	Portuguese	x	x
81.	Puerto Rico	Spanish	x	
82.	Qatar	Arabic	x	x
83.	Romania	Romanian	x	x
84.	Russia	Russian	x	x
85.	Saudi Arabia	Arabic	x	x
86.	Serbia	Serbian	x	x
87.	Singapore	English		
88.	Slovakia	Slovak	x	x
89.	Slovenia	Slovenian	x	
90.	South Africa	Afrikaans	x	
91.	South Africa	English	x	
92.	South Africa	Zulu	x	
93.	South-Korea	Korean		x
94.	Spain	Basque	x	

No	Country	Languages	Google	Nuance
95.	Spain	Catalan	x	x
96.	Spain	Galician	x	x
97.	Spain	Spanish	x	x
98.	Spain	Valencian		
99.	State of Palestine	Arabic	x	x
100.	Sweden	Swedish	x	x
101.	Switzerland	German		x
102.	Taiwan	Chinese, Mandarin (Traditional)	x	x
103.	Thailand	Thai	x	x
104.	Tunisia	Arabic	x	x
105.	Turkey	Turkish	x	
106.	Ukraine	Ukrainian	x	x
107.	United Arab Emirates	Arabic	x	x
108.	United Kingdom	English	x	x
109.	United States	English	x	x
110.	United States	Spanish	x	x
111.	Uruguay	Spanish	x	
112.	Venezuela	Spanish	x	
113.	Vietnam	Vietnamese	x	x
114.	Wales	Welsh		x

DEV

Google's speech recognition technology now has a 4.9% word error rate

EMIL PROTALINSKI @EPRO MAY 17, 2017 4:06 PM



Google CEO Sundar Pichai today announced that the company's speech recognition technology has now achieved a 4.9 percent word error rate. Put another way, Google transcribes every 20th word incorrectly. That's a big improvement from the 23 percent the company saw in 2013 and the 8 percent it shared two years ago at I/O 2015.

The tidbit was revealed at Google's I/O 2017 developer conference, where a big emphasis is on artificial intelligence. Deep learning, a type of AI, is used to achieve accurate image recognition and speech recognition. The method involves ingesting lots of data to train systems called neural networks, and then feeding new data to those systems in an attempt to make predictions.

"We've been using voice as an input across many of our products," Pichai said onstage. "That's because computers are getting much better at understanding speech. We have had significant breakthroughs, but the pace even since last year has been pretty amazing to see. Our word error rate continues to improve even in very noisy environments. This is why if you speak to Google on your phone or Google Home, we can pick up your voice accurately."

For the sake of comparison, Microsoft declared in October 2016 that it had reached speech recognition parity with humans. Its word error rate at the time was 5.9 percent, though it's not clear if the two companies are following the same standards of evaluation.

Google has been touting its speech recognition improvements for a while now. Earlier this year, the company said it had slashed its speech recognition word error rate by more than 30 percent since 2012. The main reason for the drastic improvement? Google confirmed that it's the use of neural networks.

Pichai also shared an interesting tidbit about Home's development: "When we were shipping Google Home, we were originally planning to include eight microphones... But thanks to neural networks, using a technique called 'neural beam forming', we were able to ship it with just two microphones and achieve the same quality."

So if you're surprised at how well (or poorly) Google understands what you're saying, this is why. Recognition is getting better and better, but there's still room to get that word error rate closer to 0 percent.

Google I/O 2017: Get the latest news here

VIDEO

Grid Designer's Blog

Classic Flipcard Magazine Mosaic Sidebar Snapshot Timeslide

Block Join Faceti... 2

Block Join Faceting: Intr...

Block Join Faceting in S...

Automatic Speec... 9

Scoring Join Part... 3

How to import structured...

Lucene SIMD Codec be...

Who is who in Big Data

Spark and ZooKeeper: f...

Proposing SIMD ... 1



Numeric Range Queries...

Alternative approach to ...

Segmented Filter ... 13

Grandchildren an... 4

Solr block-join su... 49



Block Join Query ... 6

Ignoring test failures at CI

Highlights from our Ope...



Solr Experience: ... 10



Solr Experience: ... 4



Spring Nested - Part III

Automatic Speech Recognition Services Comparison

Automatic Speech Recognition Services Comparison

Introduction

“Ok Google, find me a red dress.” Your long-time customer has just been invited to an important party this evening and wants to make a good impression. She’s on her way to your store right now and can’t spend any time typing in searches while she drives. Instead of saying, “Ok, Google...” wouldn’t you rather she said, “Ok, MyFavoriteStore name?”

Both Apple and Google have done a good job educating users on the value and ease of voice-controlled features. So how mature is commercial speech recognition today? As Grid Dynamics has extensive experience in eCommerce and search solutions, we decided to take a look at the current speech recognition technologies available for voice search implementation. In this article we will share the results from our experiment - comparing the quality of different speech recognition providers.

Services






Before the Experiment was started, our team reviewed multiple providers of automatic speech recognition. We have used the following criteria for selection of the service to evaluate:

- Unified, cross-platform interface. It means service availability via HTTP REST interface
- Speech recognition quality “out of the box” without any tuning for particular customer
- Free (or low price) for initial testing of service
- Speech recognition provided as a [SaaS \[https://en.wikipedia.org/wiki/Software_as_a_service\]](https://en.wikipedia.org/wiki/Software_as_a_service)

We compared the following services.

- Google
- [Nuance \[http://www.nuance.com/index.htm\]](http://www.nuance.com/index.htm)
- [AT&T \[https://developer.att.com/\]](https://developer.att.com/)

- [WIT \[https://wit.ai/\]](https://wit.ai/)
- [IBM Watson \[http://www.ibm.com/smarterplanet/us/en/ibmwatson/developercloud/\]](http://www.ibm.com/smarterplanet/us/en/ibmwatson/developercloud/)

Block Join Faceti...	2
Block Join Faceting: Intr...	
Block Join Faceting in S...	
Automatic Speec...	9
Scoring Join Part...	3
How to import structured...	
Lucene SIMD Codec be...	
Who is who in Big Data	
Spark and ZooKeeper: f...	
Proposing SIMD ...	1
 Numeric Range Queries...	
Alternative approach to ...	
Segmented Filter ...	13
Grandchildren an...	4
Solr block-join su...	49
 Block Join Query ...	6
Ignoring test failures at CI	
Highlights from our Ope...	
 Solr Experience: ...	10
 Solr Experience: ...	4
 Spring Nested - Part III	

Google

Google Speech API is not “production” ready.

- Experimental status can change API at any time
- No official API documentation or usage capabilities
- Limitations of approximately 500 requests per day, per account
- You need to join [Chromium-dev mail group \[https://groups.google.com/a/chromium.org/forum/?fromgroups#!forum/chromium-dev\]](https://groups.google.com/a/chromium.org/forum/?fromgroups#!forum/chromium-dev) and generate appropriate key in [Google developer console \[https://console.cloud.google.com/home/dashboard\]](https://console.cloud.google.com/home/dashboard)

Example of API usage:

```
curl -X POST \
--header 'Content-Type: audio/x-flac; rate=44100;' \
--data-binary @red_dress.flac \
'https://www.google.com/speech-api/v2/recognize?lang=en-us&key=
<KEY>'
```

Nuance

Nuance speech recognition REST API features:






- [Registration \[https://developer.nuance.com/public/index.php?task=register\]](https://developer.nuance.com/public/index.php?task=register) is required
- Account upgrade from Silver- to Gold-level offered for free
- Usage limitations of 5,000 requests per day

Example of API usage:

```
curl -X POST \
--header "Content-Type: audio/x-wav; codec=pcm; bit=16; rate=16000" \
--header "Accept: application/xml" \
--header "Accept-Topic: Dictation" \
--data-binary @red_dress.wav \
'https://dictation.nuancemobility.net:443/NMDPASrCmdServlet
/dictation?appId=<APP_ID>&appKey=<APP_KEY>'
```

AT&T

AT&T speech recognition REST API features:

Block Join Faceti...	2
Block Join Faceting: Intr...	
Block Join Faceting in S...	
Automatic Speec...	9
Scoring Join Part...	3
How to import structured...	
Lucene SIMD Codec be...	
Who is who in Big Data	
Spark and ZooKeeper: f...	
Proposing SIMD ...	1
 Numeric Range Queries...	
Alternative approach to ...	
Segmented Filter ...	13
Grandchildren an...	4
Solr block-join su...	49
 Block Join Query ...	6
Ignoring test failures at CI	
Highlights from our Ope...	
 Solr Experience: ...	10
 Solr Experience: ...	4
 Spring Nested - Part III	

- [Registration](https://developer.att.com/developer/flow/apiPlaygroundFlow.do?execution=e7s1) [https://developer.att.com/developer/flow/apiPlaygroundFlow.do?execution=e7s1] is required
- Required “Premium Access” payment is [\\$99/year + Usage fees](http://developer.att.com/pricing/speech-pricing-details) [http://developer.att.com/pricing/speech-pricing-details] to access automatic speech recognition
- [AT&T REST API](http://developer.att.com/apis/speech/docs) [http://developer.att.com/apis/speech/docs] uses OAuth 2.0 for authorization
- According to documentation usage limitations is [1 request per second](https://developer.att.com/support/faqs/att-developer-program-and-api-platform-faqs#what-are-maximum-transaction-rates-for-apis) [https://developer.att.com/support/faqs/att-developer-program-and-api-platform-faqs#what-are-maximum-transaction-rates-for-apis]

Example of API usage:

```
curl -X POST \
--header "Authorization: Bearer <TOKEN>" \
--header "Content-Type: audio/x-wav" \
--data-binary "@red_dress.wav" \
"https://api.att.com/speech/v3/speechToText"
```

WIT

WIT is more about [NLP](https://en.wikipedia.org/wiki/Natural_language_processing) [https://en.wikipedia.org/wiki/Natural_language_processing] (Natural Language Processing) than about plain-speech recognition.

- Main focus, besides speech recognition, is to parse out spoken phrases and extract valuable information (e.g., some voice command). The goal is to have the system “understand” voice. For example, play “Jingle Bells” when the user says, “Hi, robot! Please play me Christmas songs.”
- Github account is all that is needed to access [WIT REST API](https://wit.ai/docs/http/20141022) [https://wit.ai/docs/http/20141022]
- No account usage limitation

Example of API usage:

```
curl -X POST \
--header "Authorization: Bearer <TOKEN>" \
--header "Content-Type: audio/wav" \
--data-binary "@red_dress.wav" \
"https://api.wit.ai/speech?v=20141022"
```

IBM Watson

IBM Speech recognition REST API features:

- Public API was released to the public in early 2015

- Registration in [Bluemix \[https://console.ng.bluemix.net/\]](https://console.ng.bluemix.net/) is required
- Usage limitations of 150,000 requests per month

Example of API usage:

```
curl -X POST \
--header "Content-Type: audio/flac" \
--user <USERNAME>:<PASSWORD> \
--data-binary "@red_dress.flac" \
"https://stream.watsonplatform.net/speech-to-text/api/v1/recognize"
```

Experiment






To compare quality of speech recognition, you first need a recorded voice. As we worked on voice search features for eCommerce, we recorded eCommerce-like search phrases. We used short phrases such as: brand names, colors, sizes, etc. Here's a sample of the phrases used - "red dress," "Calvin Klein jeans" and "xl coat." We leveraged over 3,000 different phrases for this experiment and compared different conditions like gender, age and background noise (I.e., with or without noise), as well as other criteria.

We used the following sequence for experiment purposes.

- Delivered an audio file with recorded search phrases to external services
- Received recognized text from automatic speech recognition service
- Evaluated quality metrics of recognized text vs. actual search phrase

We used multiple quality metrics, such as:

- Volume of exact recognized phrases
 - Simple, but a paramount quality metric
 - Larger number of exact recognized phrases, the better quality of speech recognition results
- [Word Error Rate \[https://en.wikipedia.org/wiki/Word_error_rate\]](https://en.wikipedia.org/wiki/Word_error_rate) (WER)
 - Minimum number of words edits (I.e., insertions, deletions or substitutions) required to change one phrase into the other
 - Normalized by phrase length (basically leveraging [Levenshtein distance \[https://en.wikipedia.org/wiki/Levenshtein_distance\]](https://en.wikipedia.org/wiki/Levenshtein_distance) between two phrases working at the word level, instead of the phenomenal level)
 - Fewer number of required edits, which meant that the phrases are more like each other - offering the best quality of speech recognition

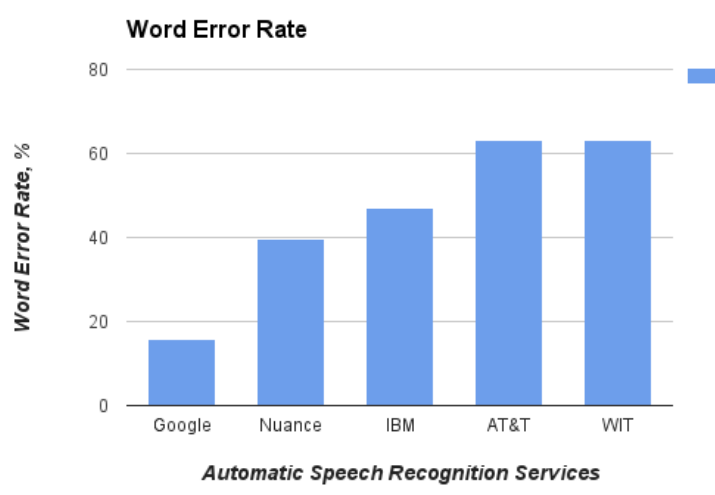
Block Join Faceti...	2
Block Join Faceting: Intr...	
Block Join Faceting in S...	
Automatic Speec...	9
Scoring Join Part...	3
How to import structured...	
Lucene SIMD Codec be...	
Who is who in Big Data	
Spark and ZooKeeper: f...	
Proposing SIMD ...	1
 Numeric Range Queries...	
Alternative approach to ...	
Segmented Filter ...	13
Grandchildren an...	4
Solr block-join su...	49
 Block Join Query ...	6
Ignoring test failures at CI	
Highlights from our Ope...	
 Solr Experience: ...	10
 Solr Experience: ...	4
 Spring Nested - Part III	

Comparison Results

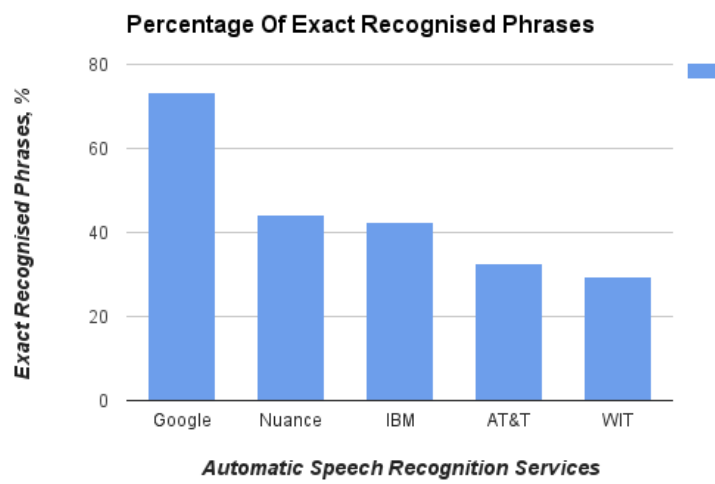
The quality champion is Google. We didn't reproduce the [declared by Google 8% WER](#) [http://venturebeat.com/2015/05/28/google-says-its-speech-recognition-technology-now-has-only-an-8-word-error-rate/] with our Grid Dynamics' data, but the results are still impressive. Google achieved 73.3% of exact recognized phrases with a 15.8% WER.

Nuance came in second place by a large margin. In Nuance, 44.1% of the phrases were recognized perfectly and the WER was 39.7%. IBM (46.9.3% and 42.3% WER) came in third place. While AT&T and WIT had the exact same WER - 63.3%, with a small advantage in exact recognition by AT&T (32.8% vs 29.5%, WIT).

Word Error Rate (less is better):



Percentage of Exact Recognized Phrases (more is better):



- Block Join Faceti... 2
- Block Join Faceting: Intr...
- Block Join Faceting in S...
- Automatic Speec... 9
- Scoring Join Part... 3
- How to import structured...
- Lucene SIMD Codec be...
- Who is who in Big Data
- Spark and ZooKeeper: f...
- Proposing SIMD ... 1
- Numeric Range Queries...
- Alternative approach to ...
- Segmented Filter ... 13
- Grandchildren an... 4
- Solr block-join su... 49
- Block Join Query ... 6
- Ignoring test failures at Cl
- Highlights from our Ope...
- Solr Experience: ... 10
- Solr Experience: ... 4
- Spring Nested - Part III

Conclusion






Based on our test criteria of exact recognized phrases and word error rate, Google is by far the best solution out of the box. This is not surprising given their history of developing and proving voice search, but unfortunately - for now - it is not commercially available. Google's quality, however, could be used as a benchmark for the commercially available products as many of them have tools and features for customizing search experience.

Exact phrase match and word error rate are only two issues to provide world-class voice search that your customers will soon expect. Additional challenges are speech recognition performance and recognizing eCommerce-specific terms. For instance, consider searches like brands, sizes, materials and, of course, long/complex phrase recognition (I.e., "Ok, MyFavoriteRetailer, find me a Ralph Lauren or Ann Taylor red cocktail dress, knee length and open back, in a size 9 that isn't dry-clean only").

But, we will discuss those challenges and our solutions in future articles.

Posted 11th January 2016 by [Andrey Kudryavtsev](#)

Labels: [API](#), [speech recognition](#), [voice search](#), [~Andrey Kudryavtsev](#)

Block Join Faceti...	2
Block Join Faceting: Intr...	
Block Join Faceting in S...	
Automatic Speec...	9
Scoring Join Part...	3
How to import structured...	
Lucene SIMD Codec be...	
Who is who in Big Data	
Spark and ZooKeeper: f...	
Proposing SIMD ...	1
 Numeric Range Queries...	
Alternative approach to ...	
Segmented Filter ...	13
Grandchildren an...	4
Solr block-join su...	49
 Block Join Query ...	6
Ignoring test failures at CI	
Highlights from our Ope...	
 Solr Experience: ...	10
 Solr Experience: ...	4
 Spring Nested - Part III	