

# **Głębokie uczenie maszynowe w systemach rozpoznawania mowy**

Paweł Łyżwa, Tomasz Woźniak



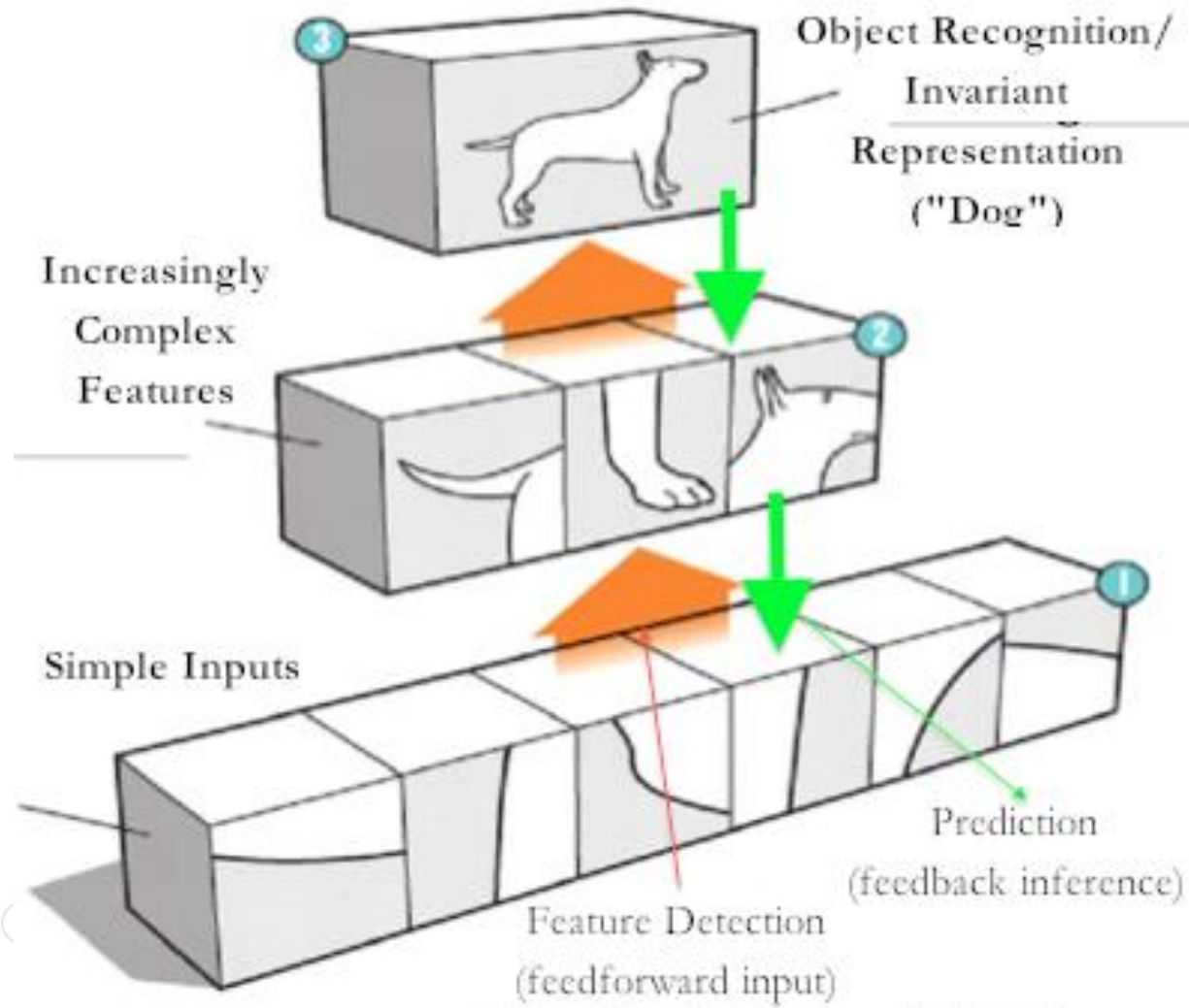
*“All recent speech recognition products by major companies either use deep neural networks — or else will soon” – George E. Dahl*

*“There is a revolution going on”*

# Czym jest głębokie uczenie maszynowe?

- ◎ nie tylko “głębokie sieci neuronowe” (DNN) ale także deep belief networks (DBN), DBN-HMM, CDNN...
- ◎ proces uczenia może być nadzorowany, nienadzorowany (najczęściej) i mieszany
- ◎ system na podstawie danych (*low-level*) uczy się rozpoznawać coraz bardziej abstrakcyjne (*higher-level*) koncepcje, tworząc warstwową strukturę

# Czym jest głębokie uczenie maszynowe?

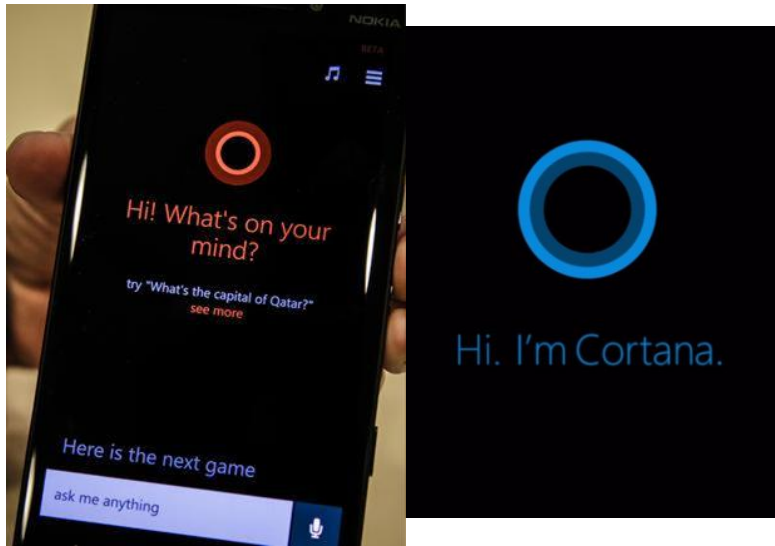


# Głębokie uczenie maszynowe - Renesans

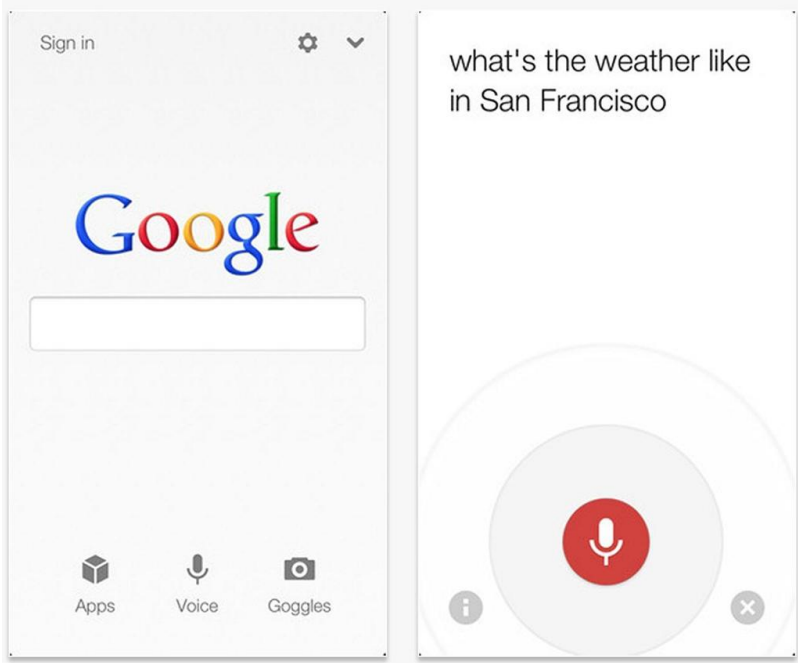
- Pomysł już w latach 80tych, jednak przez ponad 20 lat brak zadowalających efektów (HMM-GMM dawały lepsze wyniki)
- Do 2009 prawie wszystkie systemy oparte na HMM-GMM
- Co się zmieniło?
  - większa moc obliczeniowa komputerów, wykorzystanie masywnego przetwarzania równoległego -> trenowanie z miesięcy do dni
  - bardzo duże ilości danych treningowych (Bing, Google)
  - współpraca naukowców z dziedziny sztucznej inteligencji z gigantami Microsoft, Apple, Google...

# Głębokie uczenie maszynowe w rozpoznawaniu mowy

- © 2010: **Microsoft Research** publikuje osiągnięcia w rozpoznawaniu mowy z wykorzystaniem DNN - obniża dotychczasowe WER o ponad 30%
- © 2012: Narodziny **Google Brain** - początkowo system rozpoznający zdjęcia kotów i ludzkich twarzy oparty na DML (16000 rdzeni, miliard połączeń, nieanotowane dane, 3 dni treningu) - > następnie wykorzystany m.in. do rozpoznawania mowy w Android OS
- © Od 2013 : Rozpoznanie (niezakłóconej) mowy bliskie ludzkiemu, prawie wszystkie komercyjne systemy rozpoznawania mowy bazują na metodach DML



# Google Now



# Skype Translator



# "Xbox, On."

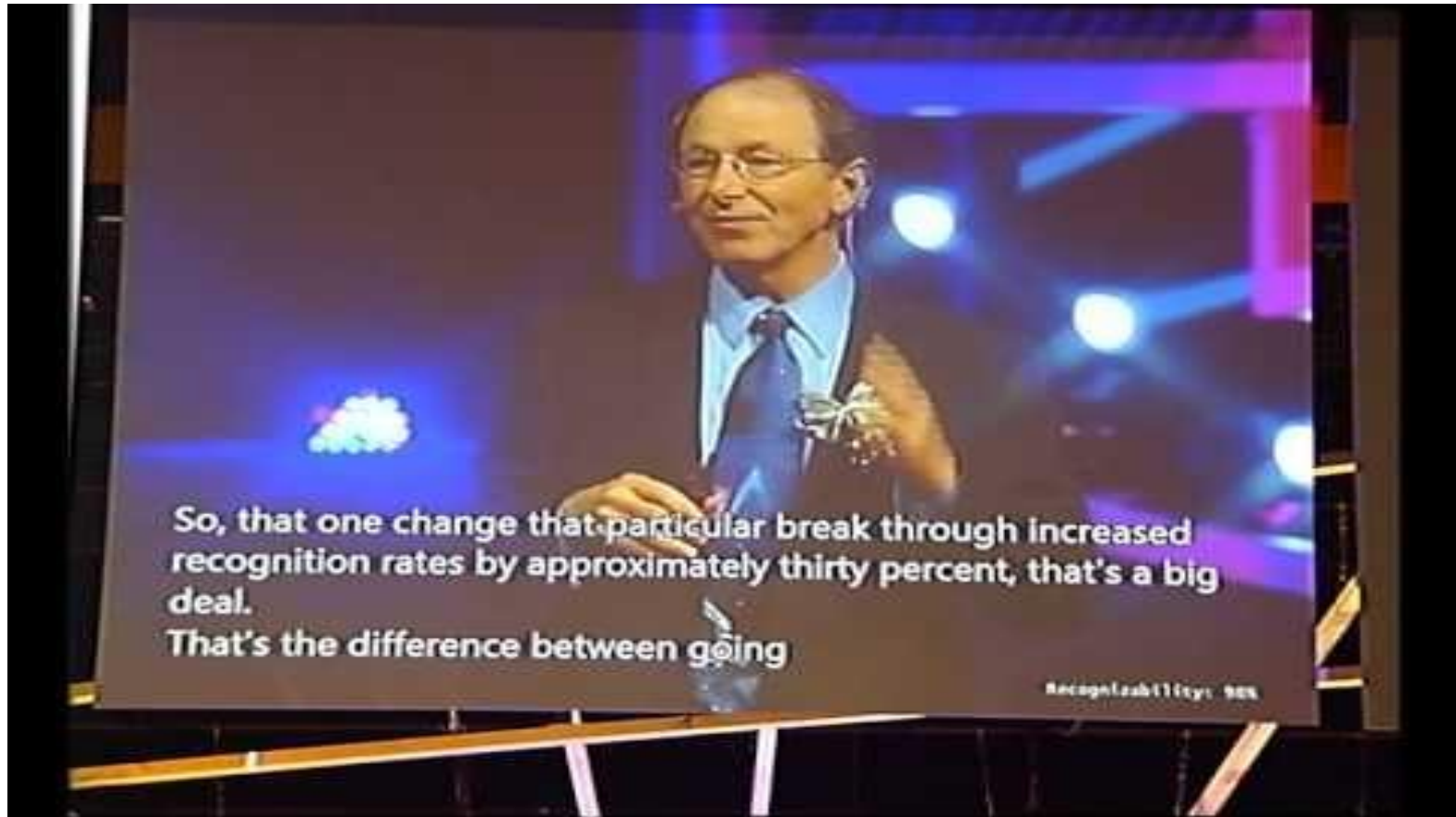




# Cortana vs Siri vs Google Now







Rok 2012, Rick Rashid (Microsoft) prezentuje system oparty na DNN który tłumaczy mówiony angielski na chiński z ER (Error-rate) < 7%

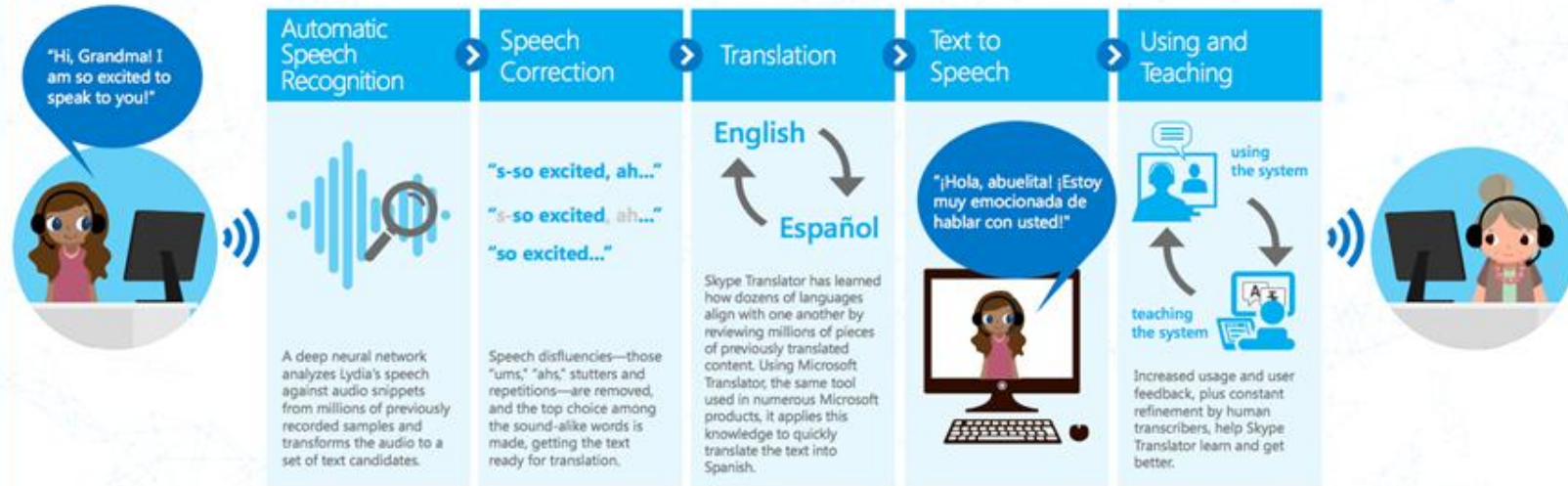
# Skype Translator

## NOW YOU'RE SPEAKING MY LANGUAGE (LITERALLY)



Skype has always been about making it easy to talk with family and friends all over the world. Now, by integrating advanced speech recognition and automatic translation into Skype, Skype Translator lets you speak with those you've always wished you could, even if they speak a different language.

### HOW SKYPE TRANSLATOR WORKS



#### PUTTING MACHINE LEARNING TO THE TEST

To provide a seamless user experience, Skype Translator uses machine learning to solve key challenges in interpreting human language, including:



Representing the different ways people really speak



Determining sentence boundaries, punctuation and case from speech

there  
they're  
their

Disambiguating sound-alike words in context



Mapping words and phrases from one language to another

### TRANSLATE INSTANT MESSAGES IN OVER 40 LANGUAGES

Holding a translated IM conversation is super easy: Choose a contact, turn on the Translation switch for that person, and start typing. When you hit enter (or tap send), your original message will appear in the right-hand pane, followed by its translation. Your contact on the other end will see something very similar, albeit with the translated message in his/her preferred language presented first. While voice translation initially supports English and Spanish only, IM translation supports over 40 languages, so feel free to experiment with them all—even Klingon!



Register for the preview at [www.skype.com/translator](http://www.skype.com/translator) and wait for your invite.

Install the Skype Translator client.

Use Skype Translator to call someone who speaks Spanish. Or, if you speak Spanish, call someone who speaks English.

Every call you make helps Skype Translator get a little bit better. You won't see the improvement right away, but you will see gradual improvement over time.





Maj 2014, Skype Translator Demo



**Dziękujemy za uwagę**

