Speech Recognition

Steve Renals

Speech Recognition Goal

systems that can detect *who* *spoke what*, *when* and *how*

for any language, acoustic environment and task domain

Speech translation of TED talks

Je suis un créateur de jeux.
J'ai fait des jeux en ligne depuis 10 ans.
Et à mon objectif pour la prochaine décennie est d'essayer de le rendre aussi facile de sauver le monde, dans la vraie vie, comme c'est de sauver le monde dans les jeux en ligne.
Maintenant "i" espère prévu pour cela, et elle comporte convaincre davantage de personnes, y compris vous tous passer plus de temps à jouer plus grands et de meilleurs jeux.
Maintenant, nous dépensons trois milliards d'heures par semaine, aux jeux en ligne.
Certains d'entre vous pourriez penser, c'est beaucoup de temps à consacrer aux jeux.
Peut-être trop de temps compte tenu, combien de problèmes urgents que nous devons résoudre dans le monde réel.
Mais en fait, selon mes recherches à l'institut pour l'avenir, c'est en fait le contraire est vrai.
Trois milliards d'heures par semaine près du jeu n'est pas assez pour résoudre les problèmes les plus urgents.
Je suis en fait "i" pense que si nous voulons survivre au siècle prochain sur cette planète, nous devons augmenter ce total de façon spectaculaire.

I'm Jane McGonigal. I'm a game designer.
I've been making games online now for 10 years,
and my goal for the next decade
is to try to make it as easy
to save the world in real life
as it is to save the world in online games.
Now, I have a plan for this,
and it entails convincing more people,
including all of you, to spend more time
playing bigger and better games.
Right now we spend three billion hours a week
playing online games.
Some of you might be thinking,
"That's a lot of time to spend playing games.
Maybe too much time, considering
how many urgent problems we have to solve in the real world."
But actually, according to my research
at The Institute For The Future,

i'm a game designer
i've been making games online out for ten years
and at my goal for the next decade is to try to make it as easy to save the world in real life as it is to save the world in online games
now i hope planned for this and it entails convincing more people including all of you to spend more time playing bigger and better games
right now we spend three billion hours a week playing online games
some of you might be thinking that's a lot of time to spend playing games
maybe too much time considering how many urgent problems we have to solve in the real world
but actually according to my research at the institute for the future it's actually the opposite is true
three billion hours a week is not nearly enough gameplay to solve the world's most urgent problems
i'm in fact i believe that if we want to survive the next century on this planet we need to increase that total dramatically
i've calculated the total we need at twenty one billion hours of gameplay every week
so that's probably a bit of a counter intuitive idea so i'll just say it again what it sink in

Distant Speech Recognition

Speech Recognition State-of-the-art

(Measured in % Word Error Rate)

- <10% on (some) lectures and TV news programmes
- <15% on (some) conversational telephone speech
- <30% on (some) multiparty conversations
- <40% on (some) TV dramas and movies

Some current projects

- Adaptation of neural network acoustic models
- Multi-task learning for NN acoustic models
- Domain adaptation
- Robustness to additive noise and reverberation in broadcast speech
- Prosodically driven recurrent neural network language models
- End-to-end speech recognition using RNNs
- Optimising RNNs

Speech Recognition Research challenges

- Fragile operation across different conditions – requires automatic adaptation to acoustic environment, speaker, genre, language …
- Degraded acoustic signals – noise, reverberation, overlapping talkers, distant microphones
- Reliance on supervised approaches
- Models include relatively little speech knowledge
- Models do not factor different causes of variability
- Systems react crudely (if at all) to the context and the environment
- Understand the acoustic scene - locate, identify, recognise all the acoustic sources

Speech recognition
Hot topics

- Learning representations using (deep) neural networks
- Learning low-dimension subspaces
- Recurrent neural networks and end-to-end systems
- Unsupervised adaptation
- Combining signal processing and machine learning
  - combining source separation and speech recognition
  - microphone array beamforming
  - direct waveform processing