Speaker Accent Classification with Deep Learning

Université de Montréal

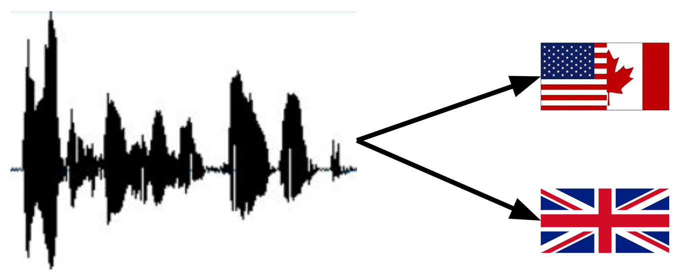
IFT 6390: Fondements de l'apprentissage machine

DIRO

Mila

Jonathan Bhimani-Burrows[1], Khalil Bibi[2], Arlie Coles[3], Akila Jeeson Daniel[4], Y. Violet Guo[5], Louis-François Préville-Ratelle[6]

The Task

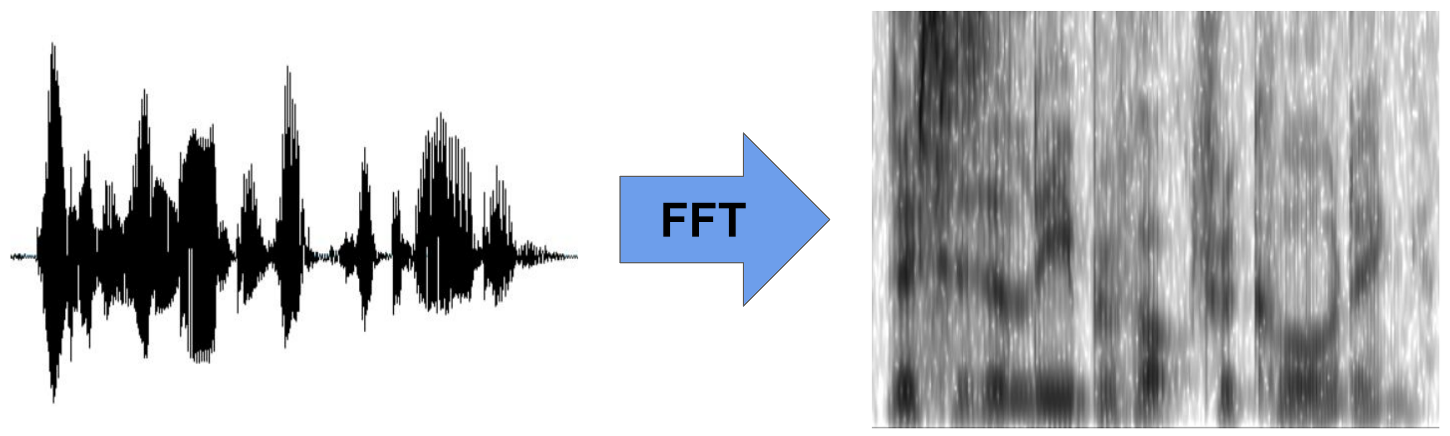


- Classify spoken audio as North American (NA) or British (UK) English.
- Useful for separating data input to ASR pipelines.

Two clean spoken corpora for training/testing:

1. **Librispeech.** NA; 5.5 hours; 40 speakers.
2. **Librit.** UK; 7 hours; 27 speakers.

Feature Extraction



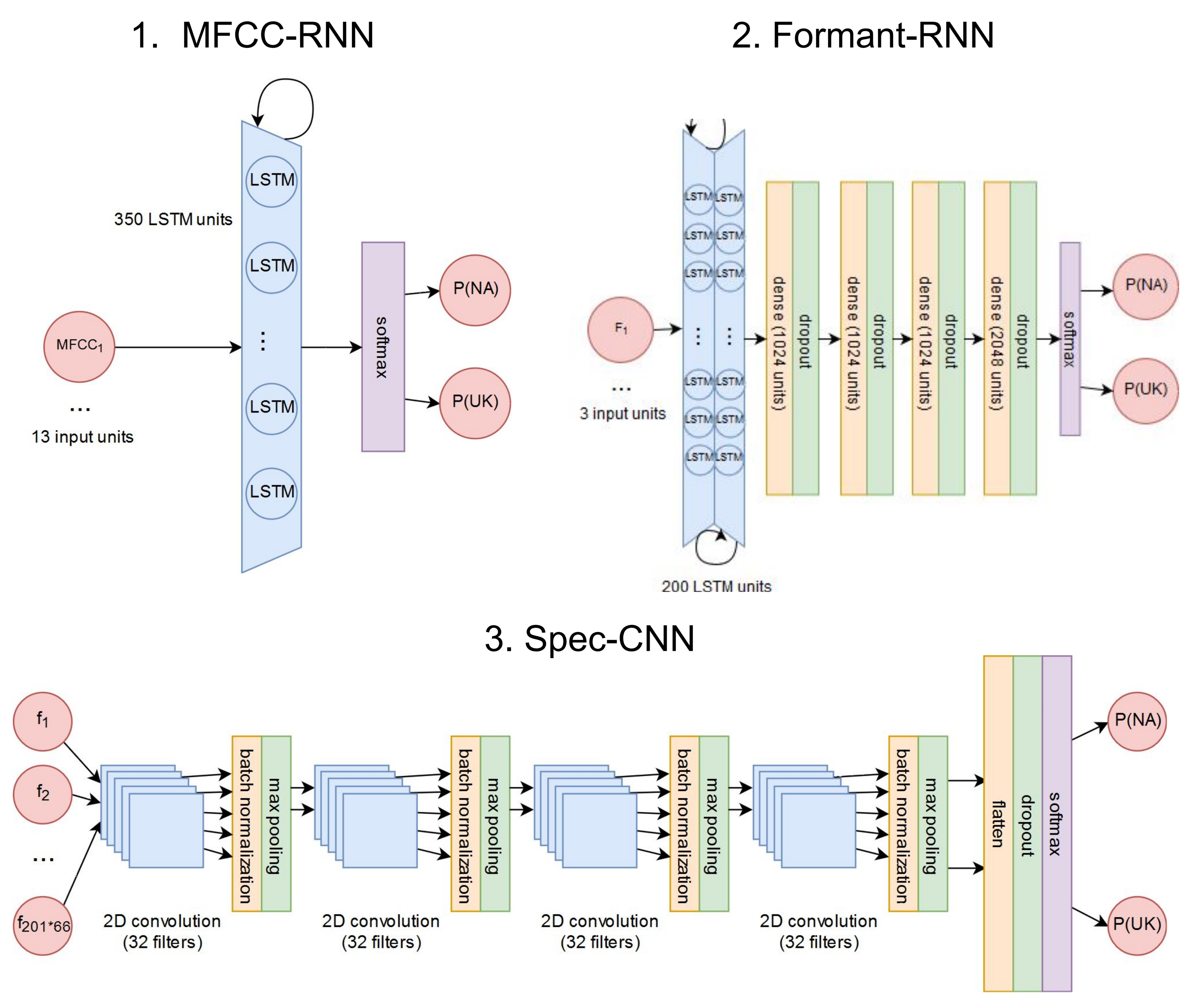
Three types of features extracted:

1. **MFCCs**. Standard in speech processing; modelled on the human ear. Apply Mel filterbank on power spectrum, then DCT.
   Result: Sequence of 13 coefficients per time slice.

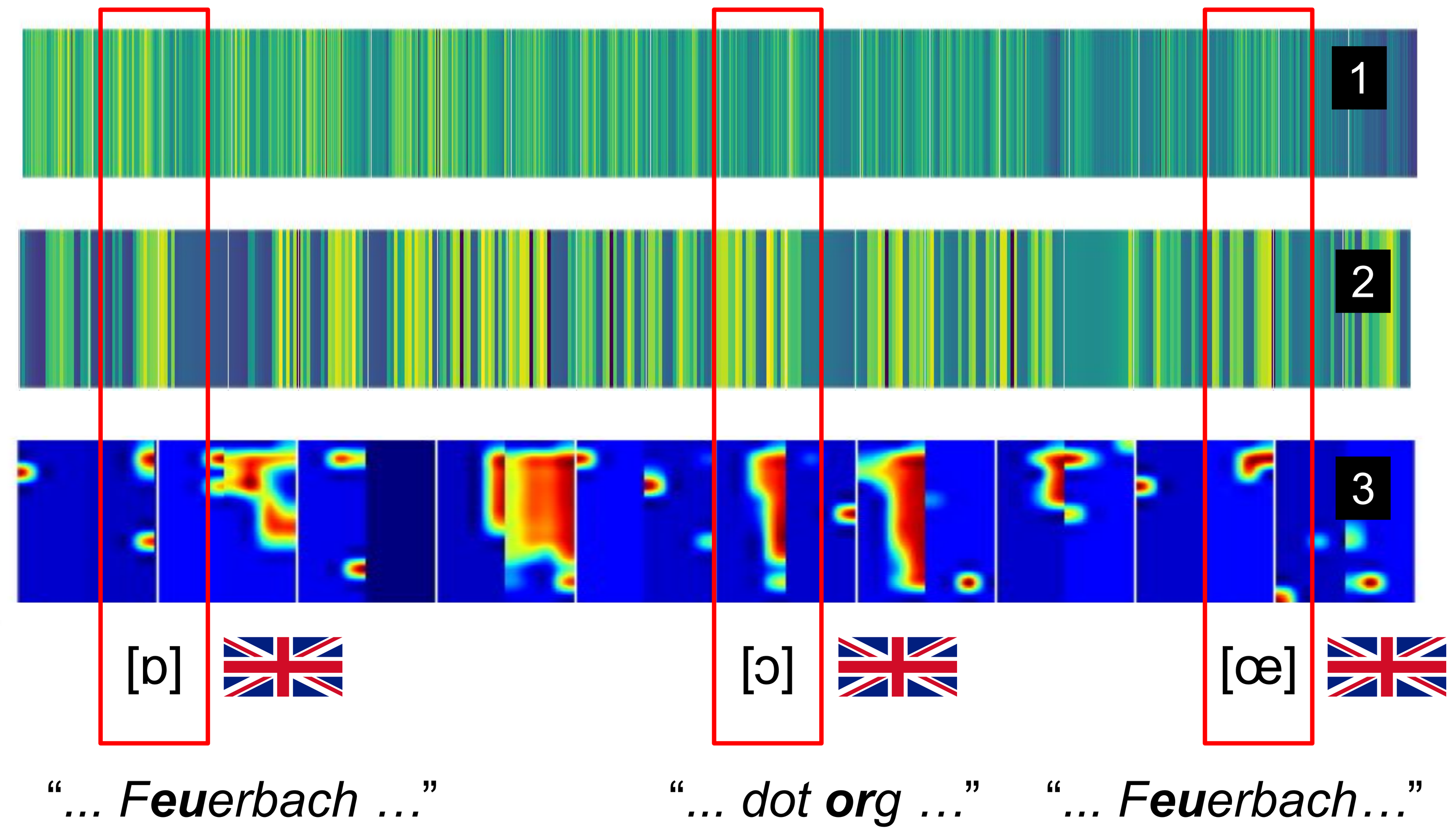
2. **Formants**. First three resonant frequencies of voiced sounds that may correlate with accent; modelled on linguistics knowledge.
   Result: Sequence of 3 frequencies per time slice.

3. **Raw spectrogram**. Frequency, amplitude, and time values after FFT needed to generate above figure; modelled on raw signals.
   Result: Concatenation of 201 amplitude-frequency pairs per time slice.

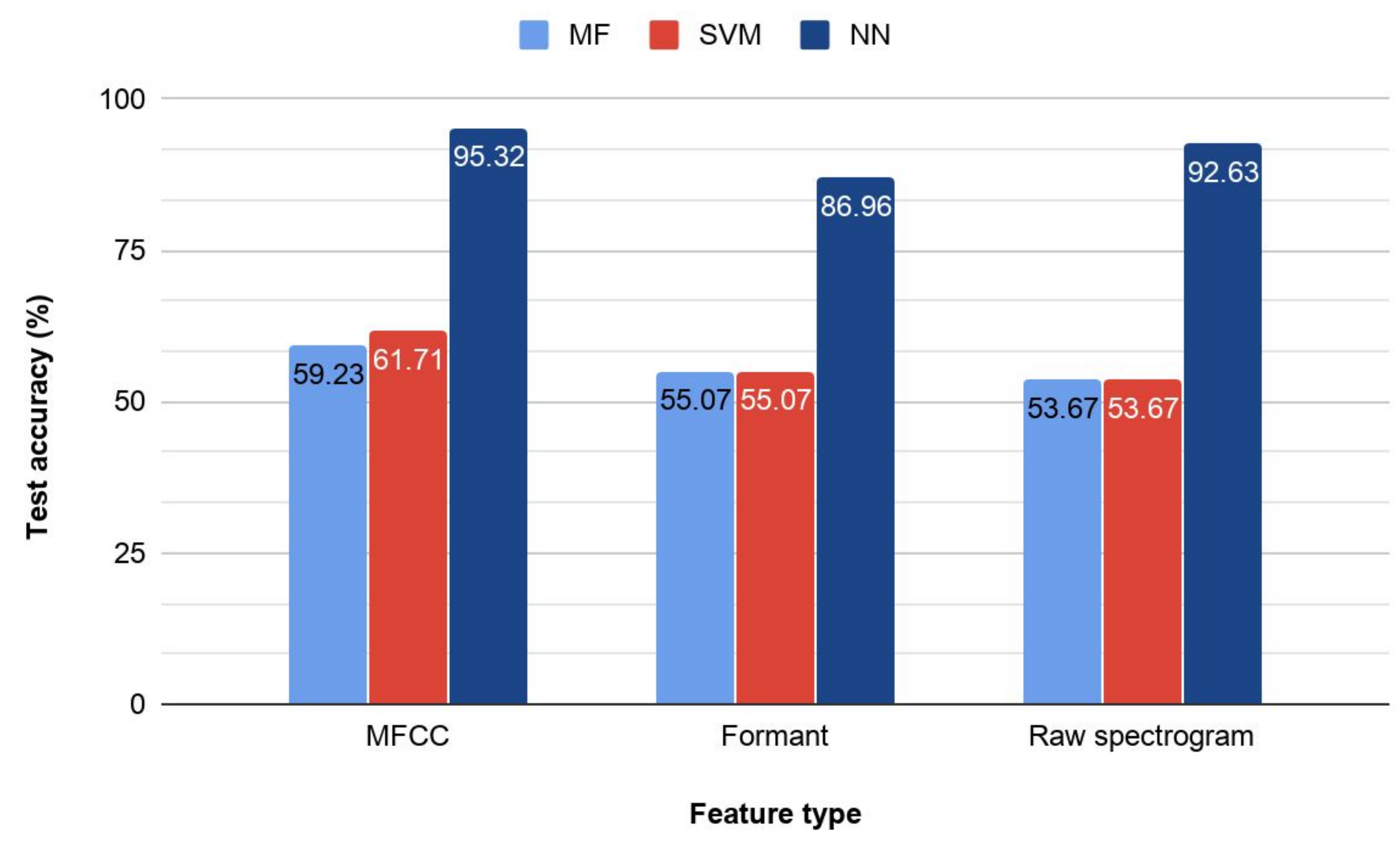
Proposed Networks



Learned Representations



Results



- All three networks outperform MF and SVM baselines.
- MFCC-RNN performs best due to context access and sufficient feature complexity.
- Spec-CNN performs well due to high feature complexity, but might be hindered by noisy signal and lack of context. More data and computational power may project long-run performance beyond MFCC-RNN.
- Formant-RNN performs most poorly due to insufficient feature complexity and limitation to voiced frames, but benefits from context access and feature interpretability.

References

Chen, Lily, Laura L Shen, and Meng Tang (2018). "Accent classification and neural style transfer of English Speech". In: *Stanford CS230 course project*, 2018. Stanford University, pp. 1–3.

Chu, Albert, Peter Lai, and Diana Le (2017). "Accent Classification of Non-Native English Speakers". In: *Stanford CS224 course project*, 2017. Stanford University, pp. 1–8.

Deshpande, Shamalee, Sharat Chikkerur, and Venu Govindaraju (2005). "Accent classification in speech". In: *Automatic Identification Advanced Technologies*, 2005. Fourth IEEE Workshop on. IEEE, pp. 139–143.

Ge, Zhenhao (2015). "Improved accent classification combining phonetic vowels with acoustic features". In: *Image and Signal Processing (CISP)*, 2015 8th International Congress on. IEEE, pp. 1204–1209.

Hansen, John HL and Levent M Arslan (1995). "Foreign accent classification using source generator based prosodic features". In: *Acoustics, Speech, and Signal Processing*, 1995. ICASSP-95., 1995 International Conference on. Vol. 1. IEEE, pp. 836–839.

Hershey, Shawn et al. (2016). "CNN Architectures for Large-Scale Audio Classification". In: *CoRR abs/1609.09430*. arXiv: 1609.09430. URL: http://arxiv.org/abs/1609.09430.

Omar, Mohamed Kamal and Jason Pelecanos (2010). "A novel approach to detecting non-native speakers and their native language". In: *Acoustics Speech and Signal Processing (ICASSP)*, 2010 IEEE International Conference on. IEEE, pp. 4398–4401.

Oord, Äaron van den et al. (2016). "WaveNet: A Generative Model for Raw Audio". In: CoRR abs/1609.03499. arXiv: 1609.03499. URL: http://arxiv.org/abs/1609.03499.

Panayotov, Vassil et al. (2015). "Librispeech: an ASR corpus based on public domain audio books". In: *Acoustics, Speech and Signal Processing (ICASSP)*, 2015 IEEE International Conference on. IEEE, pp. 5206–5210.

Ravanelli, Mirco and Yoshua Bengio (2018). "Speaker Recognition from raw waveform with SincNet". In: *CoRR abs/1808.00158*. arXiv: 1808.00158. URL: http://arxiv.org/abs/1808.00158.

[1] 20138260; jonathan.rahim.bhimani-burrows@umontreal.ca [2] 20113693; khalil.bibi@umontreal.ca [3] 20121051; arlie.coles@umontreal.ca [4] 20140681; akila.jeeson.daniel@umontreal.ca [5] 20120727; yue.guo@umontreal.ca [6] 708048; louis-francois.preville-ratelle@umontreal.ca