

Automated Assessment of Speaking


Judith Fairbairn, British Council

Dr Talia Isaacs, University of Bristol

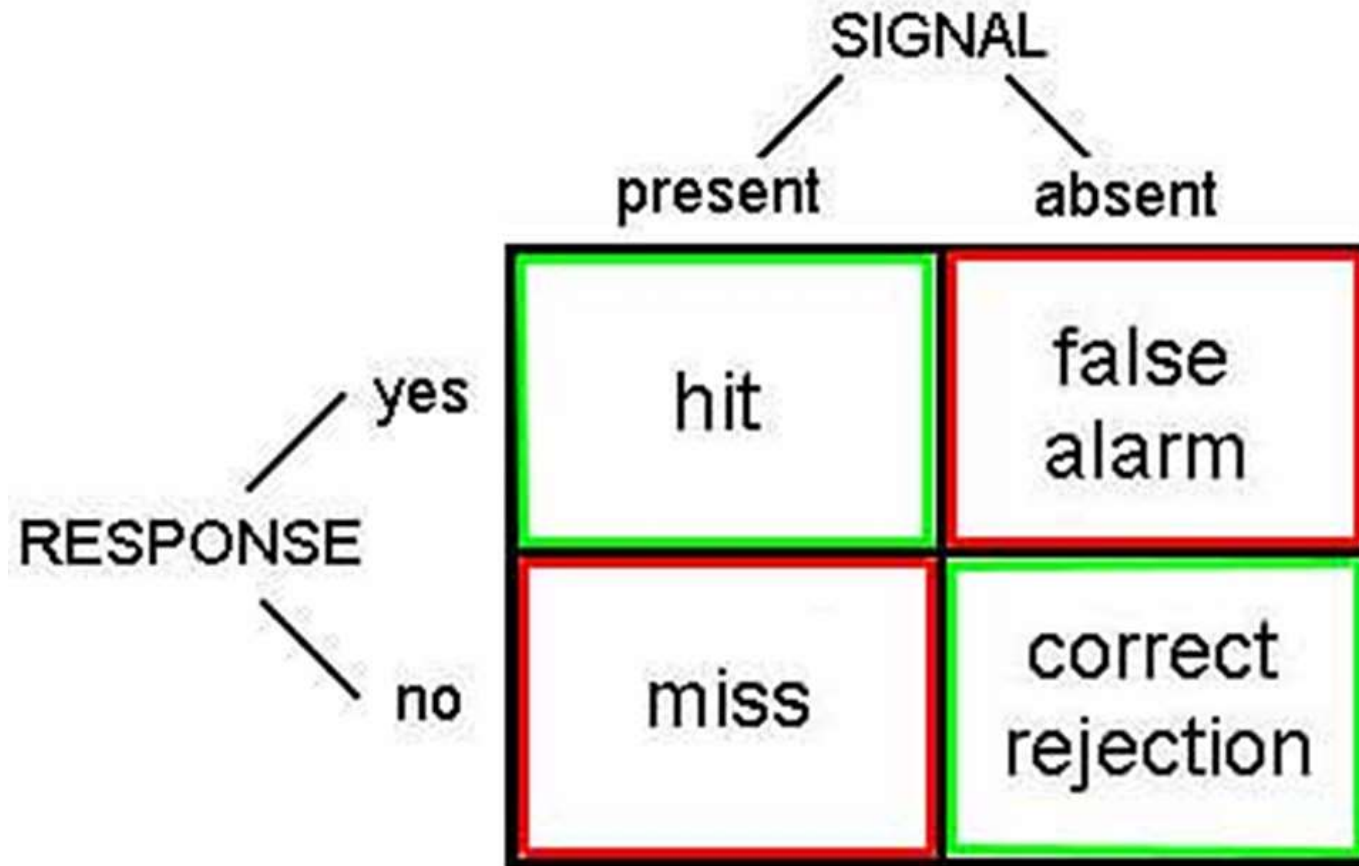
Background

- Testing speaking in school
 - Expensive
 - Time consuming
 - Lack of local expertise
- Washback effect: no test – no teaching
- Hybrid human-machine scoring model
 - Lower human rater cognitive load
 - Increase rating reliability for some linguistic features

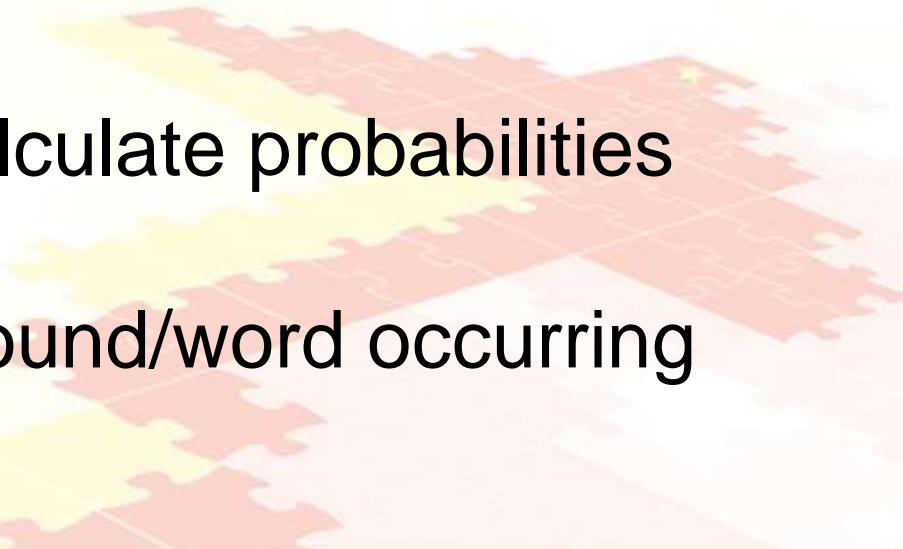
Speech recognition

- Audio recording of an acoustic signal
 - Transcription of the sounds
 - Automated assessment of writing
 - Pattern recognition using an algorithm
 - Probability of occurrence
- 

False alarms and misses



Machine training process

- Collect large volume of speech responses
 - Raters score response
 - Create training corpus
 - Statistical models calculate probabilities
 - Likelihood of each sound/word occurring
- 

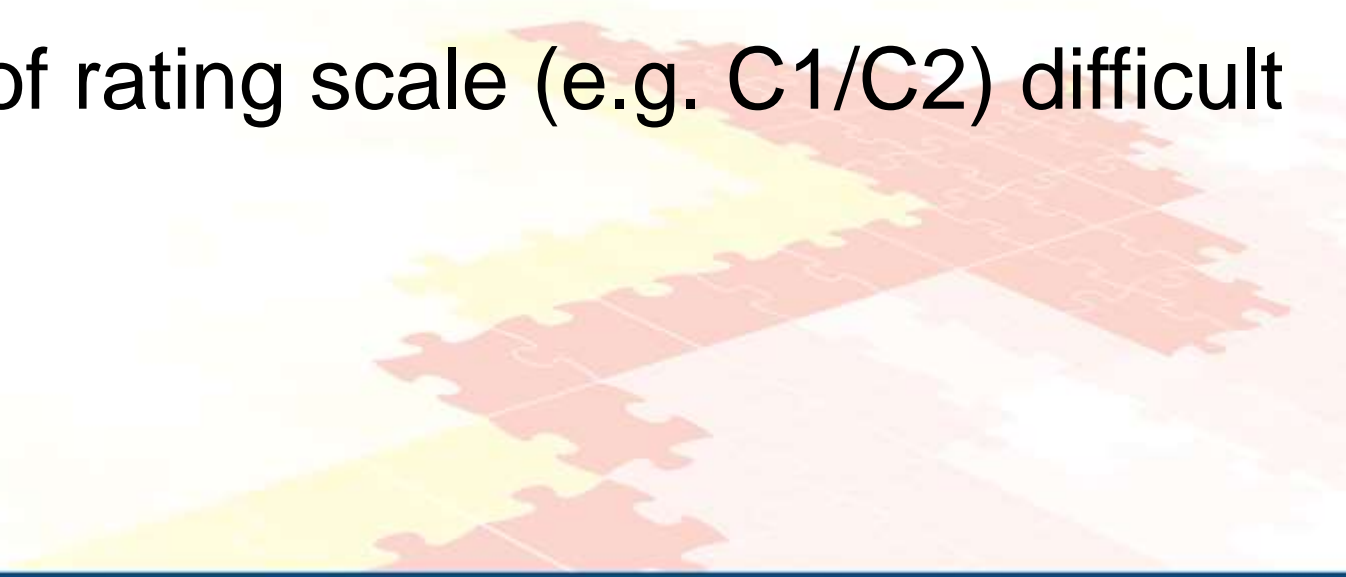
Speaking test examples

		Test mode of delivery	
		Human	Machine
Test scoring	Human	IELTS Cambridge English exams	TOEFL iBT Aptis
	Machine	NONE	Pearson PTE Pearson Versant ETS SpeechRater


Recent findings

- Individual vowel/consonant sounds and fluency easiest to measure
- Narrow construct measured (e.g. Pearson) but construct-irrelevance avoided
- Wider construct measured (e.g. ETS) but only used for low-stakes testing

Recent findings

- Correlations between computer/rater and rater/rater can be similar
 - Top end of rating scale (e.g. C1/C2) difficult to mark
- 


Building an automated test

- Task design and item writing
 - Sound quality
 - Piloting and trialling
 - Data collection
 - Annotating responses
- 

Native vs non-native pronunciation

- Local vs global
- Inner circle English vs World English
- Comprehensibility / intelligibility construct
- Expert users of Vietnamese and English model

Conclusion

- Hybrid human-machine scoring model
 - Tool to train human raters
 - Double marking
 - Vietnamese pronunciation
- 

Reference

Isaacs, T. (accepted). Fully automated speaking assessment: Changes to proficiency testing and the role of pronunciation. In O. Kang, R. I. Thomson, & J. Murphy (Eds.), *The Routledge handbook of English pronunciation*. Abingdon, UK: Routledge.

Isaacs, T., & Trofimovich, P. (2012). “Deconstructing” comprehensibility: Identifying the linguistic influences on listeners’ L2 comprehensibility ratings. *Studies in Second Language Acquisition*, 34, 475–505.

Thank you

Judith Fairbairn
Productive Skills Test Researcher
British Council

judith.fairbairn@britishcouncil.org