Elicitation and analysis of a robust word misperception corpus in Spanish

Attila Máté Tóth[1], Maria Luisa Garcia Lecumberri[1] and Martin Cooke[2,1]
(1) Language and Speech Lab, University of the Basque Country ; (2) Ikerbasque

Motivation

Speech misperceptions consistent across listeners can give valuable insights into human speech perception and can be used to refine and evaluate computational models of speech perception. Contrasting with previous work [1, 2, 3, 4] which focused on anecdotal reports of individual 'slips of the ear', we propose the laboratory elicitation of 3000+ robust Spanish word misperceptions in noise. We conduct a phonetic analysis on the confusions presented, as well as introduce a novel categorisation scheme based on the amount of information recruited from the masker present in the confused word.

Methods

Speech materials

3962 high frequency, 1-3 syllable Spanish words recorded by two male and two female talkers.

Maskers

SSN: Speech-shaped noise
BMN1: Speech modulated noise
BMN3: 3-talker babble mod. noise
BAB4: 4-talker babble
BAB8: 8-talker babble

SNR ranges were set for each of the above maskers based on [5] as well as pilot tests, and range from 1 to -4 dB for informational and -3 to -13 dB for energetic maskers.

Procedure

Adaptive techniques which prune tokens that are unlikely to lead to consistent confusions yielded a 2.6-fold increase in interesting confusion discovery rate over earlier non-adaptive techniques [5, 6].

Listeners

173 young adults (monolingual in Spanish or bilingual in Spanish/Basque) screened up to 20 blocks of 100 tokens each. A maximum of 15 listeners heard the same token.

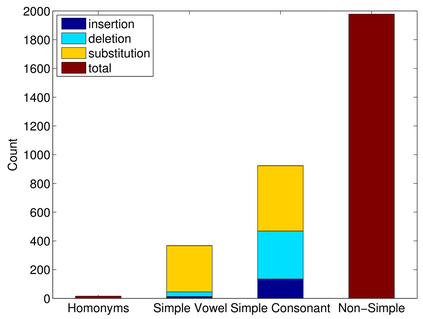
See [7] for more details on elicitation and analysis of the corpus in its initial state.

Outcome

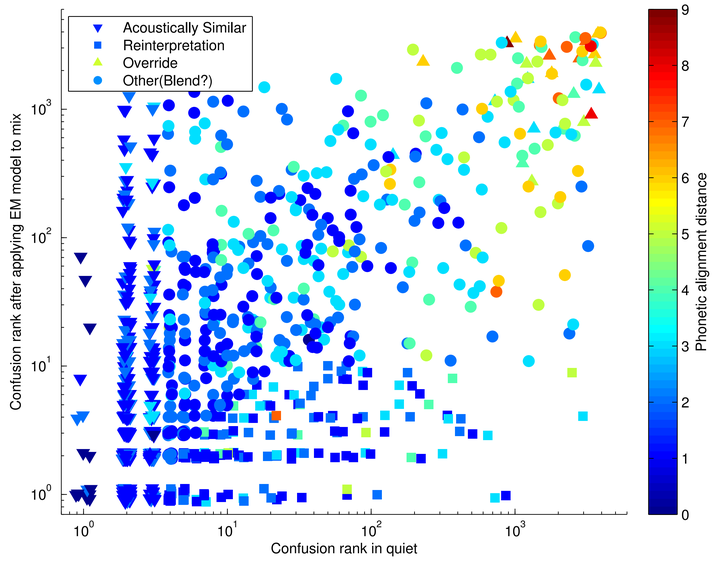
- **308 157** responses to **53 039** different tokens were collected.
- **3270** 'interesting' confusions with minimum listener agreement of 6 of 15.
- Interesting token discovery rate: **9.6** per listener hour.

Segmental errors

- The taxonomy is an extension of [1].

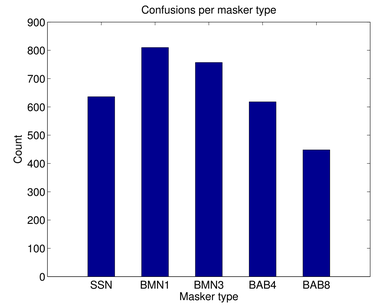


Automatic classification of confusions



- Confusions ranked in quiet (rq) and after applying EM model (rEM) [8]
- 3-state 10 mixture triphone HMMs with cochleagram representations
- Acoustically similar: $rq <= 3$
- Reinterpretation: $rEM <= 10$ & $rEM <= rq/2$
- Override: confused word can be found in babble

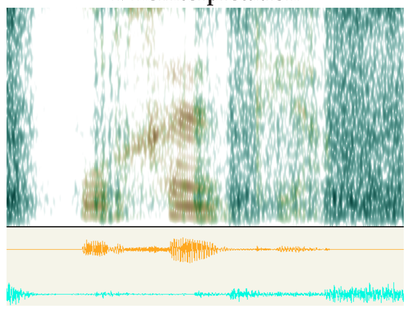
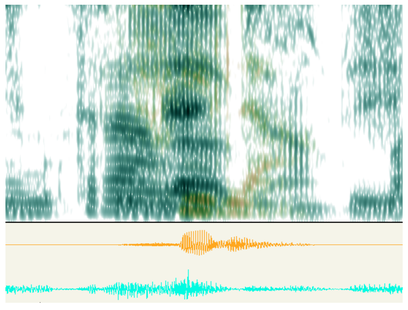
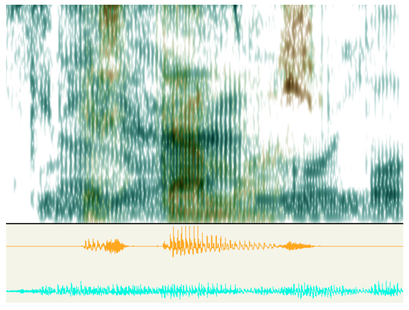
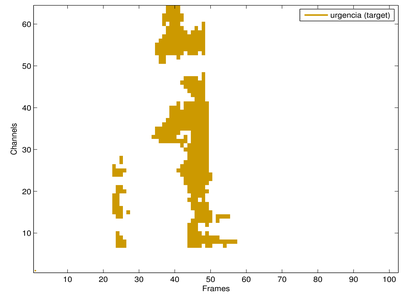
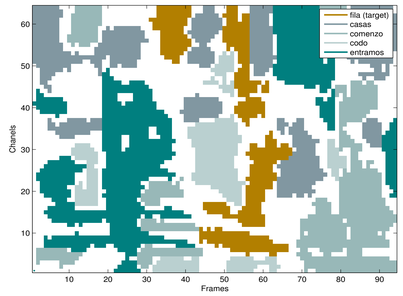
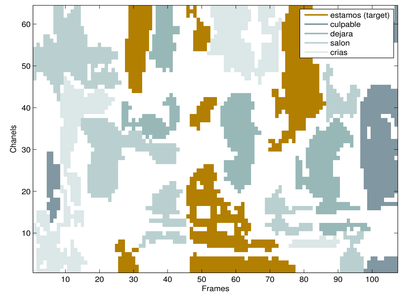
Confusions Vs. Masker



Discussion

- Microscopic perception models such as the missing data recognizer [8] and the glimpse decoder [9] can be helpful in identifying the origin of confusions.
- In turn, robust speech misperceptions help refine computational speech perception models.
- Follow-up listening tests will determine which properties of the target and masker combination lead to the misperception.
- The corpus will be released to the community as an open resource.

Speech-noise interactions: how much of masker appears in confusion?

Category	I. Reinterpretation	II. Override	III. Blend
			
			
Confusion	/urxenθia/ ↦ /muxer/ (11)	/fila/ ↦ /entramos/ (14)	/estamos/ ↦/kristal/(10)
Masker	BMN1 @ -9.6 dB	BAB4 @ -2.9 dB	BAB4 @ -0.2 dB
Other responses	s, empujar, empuje, mujre	entrar	cristales (2), quien esta, crital, estamos
Info from masker	minimal	total	partial

References

[1] S. Garnes and Z. S. Bond (1980). A slip of the ear? a snip of the ear? a slip of the year? *Errors in Linguistic Performance: Slips of the Tongue, Ear, Pen and Hand*
[2] Z. Bond (1999). Slips of the ear, *The Handbook of Speech Perception*
[3] A.Cutler and C.Henton, (2004). There's many a slip 'twixt the cup and the lip, *On Speech and Language: Studies for Sieb G. Nooteboom*
[4] K. Tang and A. Nevins, (2012). Naturalistic speech misperception - a computational corpus-based study, *Proceedings of the 43rd Meeting of the North East Linguistic Society*
[5] M. Cooke, (2009). Discovering consistent word confusions in noise, *Proc. Interspeech* pp. 1887-1890
[6] M. Cooke, J. Barker, and M. L. Garcia Lecumberri (2013). Crowdsourcing in speech perception, *Crowdsourcing for Speech Processing: Applications to Data Collection, Transcription and Assessment* pp. 141-176
[7] M. L. Garcia Lecumberri, A. M. Toth, Y. Tang, and M. Cooke (2013). Elicitation and analysis of a corpus of robust noise-induced word misperceptions in Spanish. *Proc. Interspeech* pp. 2807-2811
[8] M. Cooke, P. Green, L. Josifovski, A. Vizinho (2001). Robust automatic speech recognition with missing and unreliable acoustic data *Speech Communication* pp. 267-285
[9] J. Barker, M. Cooke, D. Ellis (2005) Decoding speech in the presence of other sources *Speech Communication* pp. 5-25

Acknowledgements

The research leading to these results was partly funded from the European Community 7th Framework Programme Marie Curie INSPIRE ITN, the Language and Speech project of the Basque Government and the Spanish Government DIACEX grant FFI 2012-31597. A special thanks to Yan Tang for software support.