Beechey, T., Buchholz, J. M., & Keidser, G. (2018). Measuring communication difficulty through effortful speech production during conversation. *Speech Communication*, *100*, 18–29. https://doi.org/10.1016/j.specom.2018.04.007

Keywords: Lombard speech, conversation, communication, effort, accommodation

What is Lombard Speech?

- Speech produced in noise
- Speech modifications to "increase the intelligibility of speech for the hearer...in difficult listening conditions"
  - Rate of vocal fold vibration, articulation of speech sounds through shape & compliance of the vocal tract, overall vocal level, rate of production, length & complexity of utterances, manner of interaction with communication partner
- Characterized by increased vocal effort – manifested in acoustic changes (ie. Increased intensity, mid-frequency emphasis, higher F1 & F0)
- Dynamic feedback
- Accommodation

Previous Lombard Speech Research

- Communicative view: speech modifications are intentionally used by talkers to increase the intelligibility of their speech for the hearer in difficult listening conditions
- Cooke & Lu (2010), Hazan & Baker (2011): measured acoustic phonetic changes in conversations, separated talkers into different booths/acoustically transparent screens – removed impt aspects of natural interaction (visual cues, sense of co-location)
- Aubanel & Cooke (2013), Aubanel et al. (2011, 2012): live competing speech as a masker – study of temporal strategies (talkers attempt to exploit predictable gaps in competing speech to maximize the intelligibility of their own speech)
- Generally, background noise is stationary or constructed babble noise
- Generally, vocal level, F0, formant frequencies, spectral tilt & vowel duration are targeted
- Little is known about how conversational dynamics affect Lombard speech

Research Question / Hypothesis

- (1) How do talkers modify their speech when communicating in realistic acoustic environments of differing complexity at both the acoustic-phonetic level and the interactive level?
- (2) What is the reliability of automated acoustic analysis? (vs. Manual annotation methods) / Can rapid acquisition of speech effort data be used in future clinical settings?
- Hypothesis: *'speech modifications at the acoustic-phonetic level (vocal level, F0, formant frequencies) will follow different patterns of change than those at the interactive level (ie. turn taking behaviour)'*
  - *'automatically extracted Lombard speech measures at multiple linguistic levels may be used to measure changes in communications difficulty & effort during conversation'*

Speaker information

- 10 male + 10 female
- Native Australian-English speaking adults
- 18-51 years old (mean = 28.7 years, SD = 7.97 years)
- No hearing impairments

Materials / Procedure

- Groupings – tested in pairs
- Sat facing one another (1.3m), wearing headset microphones
- Noise presented through headphones in 5 min blocks
- 10 experimental blocks for each pair (2 x 5 acoustic env's)
- 60s break after each block
- Headset microphone recording processed to minimize cross-talk (MATLAB)

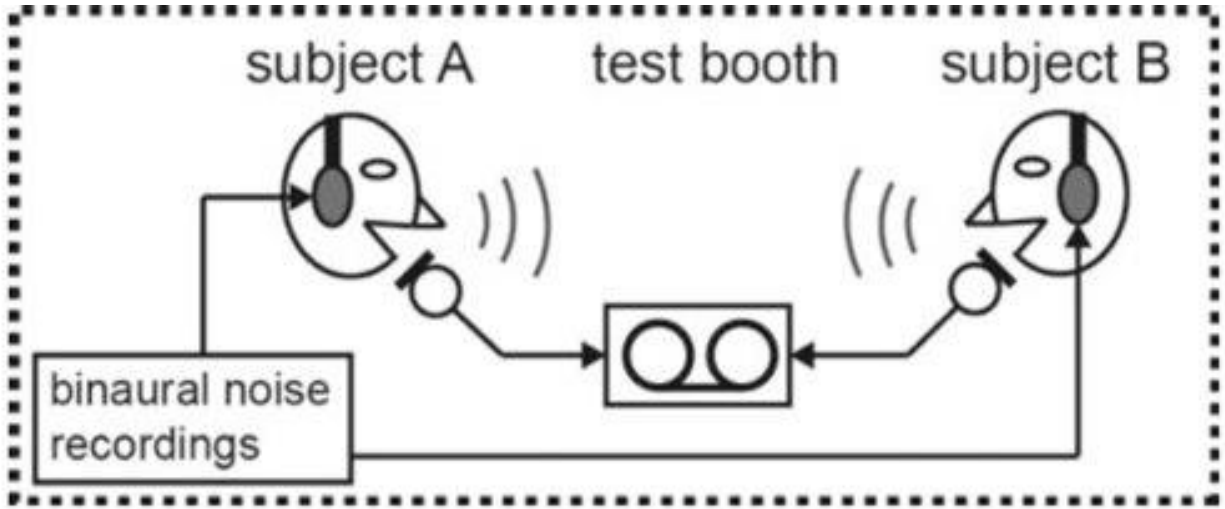
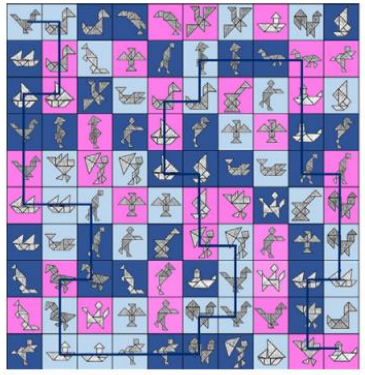
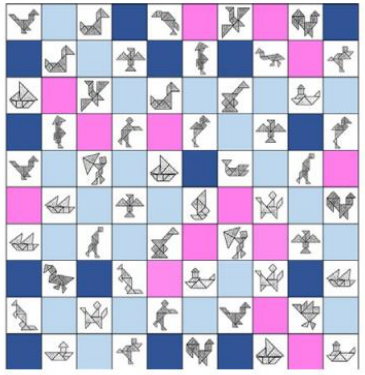


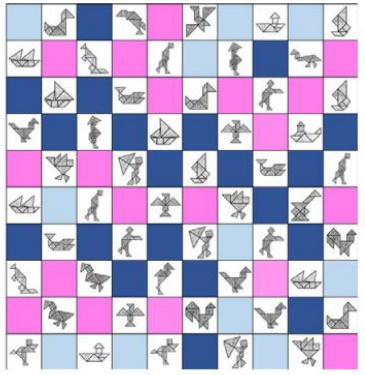
Fig. 2. Experimental set-up.

Speaker tasks

- Puzzle task: 8 puzzles available, find the unique path from start to end using identical colours or pictures, ~20 minutes to solve, displayed on a clipboard
- Purpose: to facilitate fluent, balanced conversations representative of everyday verbal communication & necessitates cooperation, communication
- 2 participants views were created – could not complete any part of the puzzle alone, r







the other

Acoustic Environments

- 5-minute experimental conditions (library, open-plan office, care, traffic on busy road, mall food court)
- 1-5 acoustic scenes in pseudo-randomized order
- Each environment played during 2 experimental blocks, total of 10 blocks per pair
- An average was produced for each measure (20 talkers x 5 environments x 2 repetition blocks)
- After each block, participants rated 'level of listening efforts' (ordinal scale of 0-12, ie. 0: 'no effort', 12: 'extreme effort')

Table 1
Acoustic environment levels and main sound sources.

Environment	Level (dB A)	Major sound sources
Library	48.5	air-conditioning
Office	56.5	soft speech, laughter, typing
Cafe	76.4	speech, coffee machine, refrigerators
Traffic	79.7	car and truck engine noise
Foodcourt	81.8	speech, music, food production

Stimuli - specific phonetic contrasts

- Middle 60s of conversations in the library (softest), café (middle), foodcourt (loudest) for all 10 pairs of participants were extracted
- Vowels **[ɐ], [ɪ], [u]** extracted from segments "d[ɐ]rk bl[u]e", "light bl[u]e", "p[ɪ]nk"
- Changes in vocal tract shape – frequencies of F1 & F2
- Changes in absorbance of the walls of the vocal tract – bandwidths of F1 & F2
- Source & filter mechanisms (amplitude of speech) - overall vocal level
- Interactive behaviour- duration of utterances & talker turn overlaps
- Analysis:
  - Via Praat – automated (F0, formant frequencies, formant bandwidths, vocal level) within 50ms windows
  - Vs. Manual analysis

Findings

- Effortful speech behaviour was grouped into 3 factors: source-filter (F0), loudness (vocal level & F1 measures), and interaction (overall utterance duration)
- Vocal level & F0 produce large effects across env's for both male & female talkers
- As conditions became more challenging, talkers increased the frequency of the voicing source of speech (F0) & modified the vocal tract
- In louder env's, talkers produced longer utterances & overlapped each others turns more often (no instances of sustained misunderstandings / communication breakdown)
- Sex-based findings:
  - Spectral changes (mid-frequency emphasis measure) were more pronounced in female speech
  - Male speech showed increasing F1 bandwidths, no change in F2 bandwidths in louder env's
  - Female speech showed no change in F1 bandwidths but decreased F2 bandwidths

Findings

- (Bottom left) Effort rated consistently higher as the SPL of the environment increased
- (Bottom right) - odds ratio calculated for each pair of adjacent acoustic env's – OR's greater than 1.0 found bw library/office, office/cafe, traffic/cafe, foodcourt/traffic
- Effect sizes: negligable-small ES for the office/library & traffic/cafe contrasts, small-medium ES for foodcourt/traffic & foodcourt/fcafe, reliably large ES for contrasts bw café/traffic/foodcourt & library/office

Table 2
Comparison of automated and manual analyses of F0, utterance duration and overlap duration.

	F0 (Hz)		Utterance duration (ms)		Overlap duration (ms)	
	Automated	Manual	Automated	Manual	Automated	Manual
Library	166.98	156.38	2644.14	1689.72	147.17	509.01
Cafe	193.42	190.49	2962.92	2244.96	161.30	617.08
Foodcourt	207.97	203.16	2953.63	2224.03	175.49	774.38

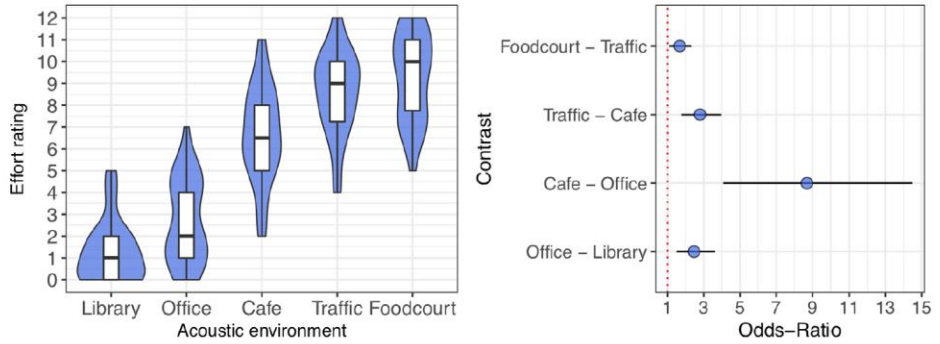


Fig. 3. Summary of subjective listening effort ratings on a 13-*point scale* (left) and odds ratios of rating listening effort in a higher category for pairs of adjacent environments (right).

Overall

- Hypothesis supported: Lombard effects at different acoustic levels are independent; intelligibility-boosting modifications occur first at the acoustic-phonetic level
- Automatic analyses found to be reliable
- Conclusion: acoustic-phonetic aspects are the most sensitive markers of effortful communication
  - Prosody, utterance length, turn-taking are also sensitive to some changes in difficult acoustic environment (more variability across talkers)