lifeLUCID = London UCL Clear Speech in Interaction Database – lifespan

Project title
Speech masking effects in speech communication across the lifespan (research project funded by the ESRC [ES/P002803/1])
https://valeriehazan.com/wp/index.php/speech-masking-effects-in-speech-communication-across-the-lifespan/

Ethical approval
The study was approved by the UCL Research Ethics Committee [Project ID number: 0534/005]. Consent forms are provided as supplementary materials.

Participants
114 monolingual native southern British English speakers (due to participant consent for usage beyond the project, 104 speakers are included here)

Participants were aged between:
- 8-12 (Young Children, CH, M=10.34 years) [CH01, CH02, CH05, CH06, CH07, CH08, CH13, CH15, CH16, CH17, CH18, CH19, CH20, CH21, CH22, CH23, CH32, CH33]
- 13-17 (Older Children, CH, M=15.94 years) [CH03, CH04, CH11, CH12, CH24, CH25, CH26, CH27, CH28, CH29, CH30, CH31]
- 18-29 years (Younger Adults, YA, M=21.82 years) [YA01, YA02, YA03, YA04, YA05, YA06, YA07, YA08, YA09, YA10, YA11, YA12, YA13, YA14, YA17, YA18]
- 30-49 years (Middle Aged, MA, M=42.98 years) [MA03, MA04, MA05, MA06, MA11, MA12, MA13, MA14, MA15, MA16, MA17, MA18, MA19, MA20, MA21, MA22, MA23, MA24]
- 50-64 (Older Middle Aged, OA, M=59.30 years) [OA01, OA02, OA05, OA06, OA07, OA08, OA26, OA27, OA32, OA33, OA39, OA40, OA41, OA42, OA43, OA44, OA45, OA46, OA47, OA48]
- 65-85 years (Older Adults, OA, M=71.19 years) [OA09, OA11, OA13, OA14, OA15, OA16, OA17, OA18, OA20, OA21, OA23, OA24, OA25, OA29, OA30, OA31, OA35, OA36, OA37]

There were 20 participants (10 female) in each of the six age bands, apart from the 13-17 band, with only 4 males. Participants were recorded in sex and age band matched pairs; they were unfamiliar with one another.

All participants were classified as normal hearing up to 4 kHz, achieving a better ear average of <25 dB across the 0.25-4 kHz octave frequencies. Participants reported no history of speech and language impairments or neurological trauma. All participants aged over 65 passed the Montreal Cognitive Assessment (MoCA; Nasreddine et al. 2005) screening test.

Procedure
Conversations were obtained while participant pairs completed a spot-the-difference puzzle in which they verbally compared two scenes, only one of which was visible to each talker (the Diapix elicitation technique; see Baker & Hazan, 2011). For these audio recordings, participants sat in adjacent acoustically-shielded rooms and communicated via headsets fitted with a cardioid microphone (Beyerdynamic DT297). They were told that they had 10 minutes to

find the 12 differences between the pictures. One of the participants (‘Speaker A’) was told to take the lead in asking questions while the other participants (‘Speaker B’) had a more passive role and primarily answered questions. Each talker participated in total of eight Diapix tasks consisting of four conditions (for four of which they were lead speaker A).

Task conditions

The four conditions were as follows:
- NORM (quiet, no masking): Participants heard each other normally.
- SPSN (speech-shaped noise): Participants communicated in speech shaped noise
- IMRE (informational masking related picture): participants communicated while three voices in the background talked about the same picture
- IMUR (information masking unrelated picture): participants communicated while three voices in the background talked about a different Diapix picture

Both IMRE and IMUR were 3-talker maskers consisting of a male, a female and a child (native British English speakers). Maskers were created using recordings obtained from previous LUCID corpora which involved participants doing the same Diapix task. The initial recordings used to create the maskers were edited to remove silences and to change the order in which the picture was described.

The intensity of all three maskers (SPSN, IMRE, IMUR) was normalised to 72 dB SPL. The intensity level of the speakers was set to approximate 0 dB SNR when speaking normally.

In order to give a more natural listening environment, we used Spatial Audio Simulation System software (Audio 3D: https://www.phon.ucl.ac.uk/resource/audio3d/) that mimics real room acoustics combined with head-related transfer functions in real-time. The maskers and the voice of the interlocutor were spatially separated by 1 metre from both each other and the “live” talker. Configuration files and the audio files for the maskers are available at: https://github.com/outepi/Diapix-virtual-room.

For each participant pair and condition, the task was run twice, with each participant switching roles as Speaker A or Speaker B. For each of these recordings, there are two textgrids and one stereo sound file that contains the speech of each of the two speakers on a separate channel. The file naming conventions are provided in a separate file (‘filenaming_conventions’).

In total therefore, this corpus includes 416 audio files and 832 textgrids.

IMPORTANT NOTE: A secondary task was carried out during Diapix conversations. At random intervals, either talker heard one out of two possible auditory cues (either a dog barking or a car horn honking) to which they had to either react by pressing a bell (dog bark) or inhibit a response (car horn). Car horn/dog barks were presented at 80 dB so +8 dB in noise conditions and 30 dB in NORM. There was a randomised order for each condition and each of the talkers (ie did not have the same sequence so they could not copy each other’s presses). The audio files therefore include dog barks and car horn honks; the textgrid files include transcriptions of <BELL> to indicate a participant's bell press.