Leeds Multi-Session Corpus of
Standard Southern British English

Manual

Contents

General information

Corpus summary

The **Leeds Multi-Session Corpus of Standard Southern British English** is a publicly available, 16-speaker dataset of Southern Standard British English speech, created at the University of Leeds between 2022 and 2024. The speakers are all females in the age range 18-35. The corpus was designed to allow for substantive analysis of intra-speaker variation. To this end, participants performed a series of commonly employed speech production and control tasks repeatedly, in 6 to 8 near-identical sessions. Consecutive sessions were separated by at least 2 weeks. Audio recordings were segmented by task to allow for the compilation of data subsets suitable for a range of phonetic analyses. The speech production tasks include a word production task, a passage reading task, and a sentence reading task in which participants produced sentences in multiple tempo and style conditions. The control tasks included a finger tapping task and a nonsense syllable repetition task. The dataset should be of interest to a range of speech researchers including those working in forensic phonetic practice.

Keywords

Speech, Southern Standard British English, Audio, Word productions, Sentence productions, Reading passage, Clear speech adjustments, Speech tempo adjustments, Intra-speaker variation.

Personnel

- Leendert Plug – Principal Investigator
- James Carter – Research Assistant
- Patricia Ternes Dallagnollo – Consultant (corpus creation, data management)
- Chris Norton – Consultant (PsychoPy, recording and audio processing)

Funding acknowledgement

The creation of this corpus was made possible by the British Academy and Leverhulme Trust Small Research Grant SG2122/210829 ‘How variable is speech, how reliable are single recordings? Assessing the medium-term dynamics of speech through iterative data collection’, awarded to Leendert Plug.

Ethics and data management

The informed consent and participant payment methods were approved by the University of Leeds (Faculty of Arts, Humanities and Cultures) Research Ethics Committee (LTSLCS-155). The data management plan was scrutinised by Research Data Leeds (RDL #337252) to ensure compliance with General Data Protection Regulation.

Citation

When using LMS materials in published research, please use the following citation:

- Plug, L. (2024), Leeds Multi-Session Corpus of Standard Southern British English [Dataset], https://doi.org/10.5518/1547.

Corpus structure

Directory hierarchy

The materials are organised in directories as follows.

Directory	Content	Sub-directories
`Annotations`	text transcripts (.txt); PRAAT TextGrids (.TextGrid)	by task
`Data_...`	audio files (.wav), data frames (.csv)	by task
`Doc`	corpus documentation (.pdf)	none
`Meta`	participant and session information	none
`Tools`	programming files with guidance	none

By-task sub-directory naming in `Annotations` and `Data_...` is as follows. (In the case of `Data_...` the 'sub-directories' are listed at top level to keep download sizes manageable: `Data_LMS_PAS`, `Data_LMS_SCL` and so on.) Data format is task-dependent, and annotation files only exist for tasks which generated audio files.

Sub-directory	Task	Data format	Annotation files
`LMS_PAS`	Passage reading	.wav	yes
`LMS_SCL`	Sentence reading (clear)	.wav	yes
`LMS_SFA`	Sentence reading (fast)	.wav	yes
`LMS_SNO`	Sentence reading (normal)	.wav	yes
`LMS_SYL`	Syllable repetition	.wav	yes
`LMS_TAP`	Finger tapping	.csv	no
`LMS_WOR`	Word reading	.wav	yes

Filenames

Within the by-task sub-directories in `Annotations` and `Data`, files are named as follows:

`LMS_`	`WOR_`	`01_`	`F_`	`S1`
corpus	task	participant number	participant gender	session number

Task labels match the sub-directory names above (PAS, SCL, SFA, SLI, SNO, SYL, TAP, WOR). Participant numbers run from 01 to 16. Participant gender is F only. Session numbers run from S1 to S8.

Text transcripts are provided in two versions: the transcript which the participant was instructed to produce ('target' transcript: `_TRG`), and the transcript which the participant actually produced ('produced' transcript: `_PRD`). The two transcripts diverge if a participant made production errors that were not corrected in the session. The 'produced' transcript formed the input for forced alignment.

Missing files

The .wav, .TextGrid and .txt files associated with the task recording LMS_SCL_14_F_S4 are known to be missing due to processing error.

Participants

Recruitment method

We set out to recruit up to 20 speakers with similar accents of English who would be able to attend a series of up to 8 recording sessions at the University of Leeds. To this end, we disseminated a recruitment call through modules taught at the University of Leeds and Leeds Beckett University and social media channels. The call asked potential participants to express interest, declare that they met the eligibility criteria and provide several items of demographic information through a linked Microsoft Form.

The full eligibility criteria were:

- native UK English speaker
- no known speaking or hearing impairment
- between 18 and 35 years old

The elicited items of demographic information were:

- gender identity ('woman', 'man', 'transgender', 'non-binary' or 'I prefer not to say')
- county in which the participant grew up (free text response)
- linguistic environment in which the participant grew up ('monolingual household' or 'multilingual household')
- own description of accent (free text response)

A copy of the Microsoft Form can be found in the `Doc` directory.

Demographics

Based on the responses to the Microsoft Form we established that we were most likely to reach our maximum recruitment of 20, while keeping the participant pool maximally homogeneous, by proceeding with female ('woman') speakers of Standard Southern British English ('Southern', 'Standard southern English', 'Southern British', 'Southern and a bit posh') who grew up in monolingual households in Sussex, Surrey, Hampshire, Kent or Buckinghamshire. We disseminated a second recruitment call with adjusted eligibility criteria.

A spreadsheet with participant demographics can be found in the `Meta` directory.

Initial meeting

Participants attended an initial meeting in which the Research Assistant explained the nature of the tasks that would make up the recording sessions; participants signed the consent form; and the Research Assistant and participant planned a schedule of sessions over the following months such that consecutive sessions were separated by at least two weeks.

A blank copy of the consent form can be found in the `Doc` directory. A spreadsheet with session information, including session dates, can be found in the `Meta` directory.

Reimbursement

Participants were told in the recruitment call and initial meeting they would be paid £80 after completing 8 sessions. Participants who completed fewer sessions, including participants who withdrew from the study, were paid £10 per completed session. Participants were paid in cash or e-voucher.

Task design

Affective slider

Participants were asked to self-report how happy and how alert they felt. We used the 'affective slider' (Betella and Verschure, 2016) to elicit this information in quantitative terms. The affective slider responses can be found in the session information spreadsheet in the `Meta` directory.

Finger tapping

Participants were asked to tap their fingers on a laptop touchpad for 20 seconds, at a pace that felt comfortable. A version of this task is used by Lidji et al. (2011), among others.

Syllable repetition

Participants were asked to produce the syllable /pa/ repeatedly for 20 seconds, if possible on a single outbreath, at a pace that felt comfortable. A version of this task is used by Ruspantini et al. (2012), among others.

Word reading

Participants were asked to read a series of words in the carrier sentence 'Now say … for me', at a pace that felt comfortable. Participants were asked to repeat the entire sentence if they made a production error. Participants produced each word twice.

The word list is based on that of the *Illustrations of the IPA* for varieties of English. It is particularly close to that used by Cox and Palethorpe (2007), with edits to make the list as uniform as possible in terms of vowel context for consonants and consonant context for vowels.

Structure	Cai#	Cai/s,z/	/b,g/Vd
Words	pie, buy, tie, die, Kai, Guy, fie, vie, sigh, thigh, thy, shy, high, my, nigh, rye, lie, why	nice, nighs, dice, dies, lice, lies	bead, bid, bed, bad, bud, bod, good, board, booed, barred, bird, bayed, bide, Boyd, bode, bowed, beard, bared

Passage reading

Participants were asked to read *The North Wind and the Sun*, the reading passage used in most *Illustrations of the IPA* published by the International Phonetic Association (Baird et al., 2022). Participants were asked to read the whole passage first, before producing it. They were given no specific instructions regarding speech tempo or style.

> The North Wind and the Sun were disputing which was the stronger, when a traveller came along wrapped in a warm cloak. They agreed that the one who first succeeded in making the traveller take his cloak off should be considered stronger than the other. Then the North Wind blew as hard as he could, but the more he blew the more closely did the traveller fold his cloak around him, and at last the North Wind gave up the attempt. Then the Sun shone out warmly, and immediately the traveller took off his cloak. And so the North Wind was obliged to confess that the Sun was the stronger of the two.

Sentence reading

Participants were asked to read a series of short declarative sentences. 15 sentences were selected from the set of 'news-like' sentences used by Smiljanić and Bradlow (2008), among others. Participants produced each sentence twice in each of three conditions:

1. 'Normal' – Participants were asked to produce the sentences at a pace that felt comfortable.
2. 'Fast' – Participants were asked to produce the sentences at what they considered a fast tempo. They were instructed that the tempo should still allow them to produce each sentence fluently and intelligibly, and that they should repeat the entire sentence if they made a production error. A similar task is used by Dellwo et al. (2005), among others.
3. 'Clear' – Participants were asked to produce the sentences as clearly as they could, as they might do if they were talking to someone in a noisy environment or to someone who has hearing loss. They were encouraged to exaggerate the movements of their mouths and were told that their speech would probably be slower and louder than normal. A similar task is used by Smiljanić and Bradlow (2008), among others; the instructions are closest to those of Tjaden et al. (2014).

1. The local train left the station more than five minutes ago.
2. The next local elections will take place during the winter.
3. Much more money will be needed to make this project succeed.
4. The art gallery in this street was opened only last week.
5. In this famous coffee shop you will eat the best doughnuts in town.
6. The last concert given at the opera was a tremendous success.
7. The committee will meet this afternoon for a special debate.
8. This rugby season promises to be a very exciting one.
9. Finding a job is difficult in the present economic climate.
10. Trade unions have lost a lot of their influence during the past ten years.
11. Seven paintings of great value have recently been stolen from the museum.
12. This is the first time an international exhibition takes place in this town.
13. The rebuilding of the city started the very first day after the earthquake.
14. The latest events have caused an outcry in the international community.
15. The recent rainfall has caused very severe damage in the higher valleys.

Picture naming

Between the main tasks, participants were asked to name or describe each of 3 greyscale line drawings which appeared on screen sequentially. This was to create a short break from reading with a controlled alternative activity. The line drawings were taken from the MultiPic database (Duñabeitia et al., 2018). The complete collection of 750 drawings can be found in `Tools\LMS_PsychoPy\test_images`.

Participants' responses for the picture naming task are not included in the corpus.

Recording procedure

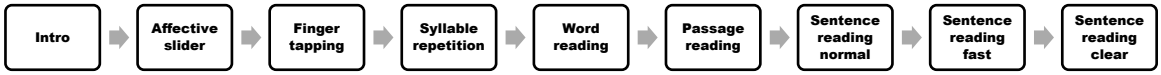
Location and equipment

All recording sessions took place in the same recording studio at the University of Leeds. The studio has a recording space and a small control room separated by a door and large

window. The recording space lay-out was kept as constant as possible from session to session. Participants were seated at a table in front of a laptop computer. A single cardioid condenser microphone (Audio-Technica AT2020) was set up approximately 30 centimetres away from the participant's mouth using a microphone stand, with pop shield. The microphone signal was recorded in the control room via a USB audio interface (M-Audio Fast Track Pro) and the software Audacity (Audacity Team, 2021) running on a Windows PC. The recordings were produced at a sampling rate of 44100 Hz with 32-bit amplitude resolution, then exported as mono WAV files with 16-bit resolution. The Research Assistant was seated in the control room, monitoring the participant's progress via headphones.

Interface

The on-screen interface was coded in PsychoPy (Peirce, 2009) for running in version 2022.2.4. The interface files can be downloaded from `Tools\LMS_PsychoPy`. The general experiment flow was as follows. At the place of each arrow, participants did a brief distractor task (picture naming).



In the introductory on-screen instructions, participants were told that the session would comprise multiple tasks, and that they should read the instructions before each task carefully. The crucial instruction wordings for individual tasks are included in the task design section of this manual. In the reading tasks, participants hit the space bar to proceed to a next word or sentence. The finger tapping and syllable repetition tasks had fixed timing (20 seconds). In the picture naming task, each drawing was displayed for 5 seconds. Participants' Affective slider and Finger tapping responses were extracted into separate data frames from the PsychoPy response file. All other responses were recorded as audio.

Randomisation

The general experiment flow was constant from session to session: that is, participants performed the tasks in the same order every time. However, within tasks there was some degree of randomisation. Where the order of task items was randomised, the items were presented in a unique random order in each individual recording session.

Task	Randomisation
Affective slider	Slider order
Finger tapping	None
Syllable repetition	None
Word reading	Word order (across 2 instances of each word)
Passage reading	None
Sentence reading (normal)	Sentence order (across 2 instances of each sentence)
Sentence reading (fast)	Sentence order (across 2 instances of each sentence)
Sentence reading (clear)	Sentence order (across 2 instances of each sentence)
Picture naming	Random sample (from 750 drawings) on each iteration

Audio processing

Manual segmentation

Once audio was recorded, initial chunking was conducted manually in Audacity (Audacity Team, 2012). The six speaking tasks were delimited and corresponding audio exported as separate WAV files. Hesitations, false starts, error corrections, interaction with the research assistant, and coughing or other nonverbal utterances were removed where they could be delimited between sentence productions.

Forced alignment

For the tasks PAS, SCL, SFA, SNO and WOR, we created annotation files with phone-level segmentation using WebMAUS (Kisler et al., 2017).

As explained in the section on filenames above, we generated two text transcripts for each task: one which contained all text which the participant was instructed to produce ('target' transcript: `_TRG`), and one which contained all text which the participant actually produced ('produced' transcript: `_PRD`). The latter formed the input for forced alignment. The two versions diverge if the participant made a production error which was not corrected in the session. Such errors were mainly of the following types.

1. Deletion — In these instances, words were omitted from a task's `_PRD` file to reflect the produced realisation. For example:

`_PRD`	The local elections will take place during the winter.
`_TRG`	The next local elections will take place during the winter.

2. Insertion — In these instances, words were added to a task's `_PRD` file to reflect the produced realisation. For example:

`_PRD`	The committee will meet this afternoon for a very special debate.
`_TRG`	The committee will meet this afternoon for a special debate.

3. Word replacement — In these instances, words were changed in a task's `_PRD` file to reflect the produced realisation. For example:

`_PRD`	Now say barred for me.
`_TRG`	Now say bared for me.

4. Repetition, hesitation and false starts, where produced sentence-internally and not corrected — These instances were transcribed as accurately as possible in a task's `_PRD` file. Transcriptions are generally recognisable by a dash. For example:

`_PRD`	They agreed that the one who first succeeded in making the traveller take his cloak off should be considered the stong-stronger than the other.
`_TRG`	They agreed that the one who first succeeded in making the traveller take his cloak off should be considered stronger than the other.

Annotation files were generated by dropping audio and corresponding `_PRD` text transcript files into the WebMAUS Basic interface (https://clarin.phonetik.uni-muenchen.de/BASWebServices/interface/WebMAUSBasic). The language model was

'English (GB)'; output format 'Praat (TextGrid)'. The resulting TextGrids, for use in Praat (Boersma and Weenink, 2023), contain three tiers:

1. ORT-MAU, containing word-level segmentation with orthographic transcription
2. KAN-MAU, containing word-level segmentation with X-SAMPA canonical form transcription
3. MAU, containing phone-level segmentation with X-SAMPA surface form transcription

TextGrids are provided as produced by WebMAUS, so will need manual checking if used for research purposes. For X-SAMPA symbol mappings, see https://www.phon.ucl.ac.uk/home/sampa/

For the task SYL, segmentation was done manually. Each TextGrid contains a point tier with boundaries placed just before the /p/ bursts. This allows for inter-burst interval durations to be extracted.

References

AUDACITY TEAM 2021. Audacity®: Free audio editor and recorder. 3.0.0 ed. https://audacityteam.org/.

BAIRD, L., EVANS, N. & GREENHILL, S. J. 2022. Blowing in the wind: Using 'North Wind and the Sun' texts to sample phoneme inventories. *Journal of the International Phonetic Association,* 52**,** 453-494.

BETELLA, A. & VERSCHURE, P. F. M. J. 2016. The Affective Slider: A Digital Self-Assessment Scale for the Measurement of Human Emotions. *PLOS ONE,* 11**,** e0148037.

BOERSMA, P. & WEENINK, D. 2023. Praat: Doing phonetics by computer. 6.3.10 ed. www.praat.org.

COX, F. & PALETHORPE, S. 2007. Australian English. *Journal of the International Phonetic Association,* 37**,** 341-350.

DELLWO, V., STEINER, I., ASCHENBERNER, B., DANKOVIČOVÁ, J. & WAGNER, P. The BonnTempo-Corpus and BonnTempo-Tools: A database for the study of speech rhythm and rate. Proceedings of Interspeech 2005, 2005 Lisbon.

DUÑABEITIA, J. A., CREPALDI, D., MEYER, A. S., NEW, B., PLIATSIKAS, C., SMOLKA, E. & BRYSBAERT, M. 2018. MultiPic: A standardized set of 750 drawings with norms for six European languages. *Quarterly Journal of Experimental Psychology,* 71**,** 808-816.

KISLER, T., REICHEL, U. D. & SCHIEL, F. 2017. Multilingual processing of speech via web services. *Computer Speech & Language,* 45**,** 326–347.

LIDJI, P., PALMER, C., PERETZ, I. & MORNINGSTAR, M. 2011. Listeners feel the beat: Entrainment to English and French speech rhythms. *Psychonomic Bulletin & Review,* 18**,** 1035-1041.

PEIRCE, J. 2009. Generating stimuli for neuroscience using PsychoPy. *Frontiers in Neuroinformatics,* 2.

RUSPANTINI, I., SAARINEN, T., BELARDINELLI, P., JALAVA, A., PARVIAINEN, T., KUJALA, J. & SALMELIN, R. 2012. Corticomuscular coherence is tuned to the spontaneous rhythmicity of speech at 2-3 Hz. *Journal of Neuroscience,* 32**,** 3786-3790.

SMILJANIĆ, R. & BRADLOW, A. R. 2008. Temporal organization of English clear and conversational speech. *The Journal of the Acoustical Society of America,* 124**,** 3171-3182.

TJADEN, K., SUSSMAN, J. E. & WILDING, G. E. 2014. Impact of Clear, Loud, and Slow Speech on Scaled Intelligibility and Speech Severity in Parkinson's Disease and Multiple Sclerosis. *Journal of Speech, Language, and Hearing Research,* 57**,** 779-792.