

Montclair Map Task Corpus of Conversations in English

Authors: Jennifer S. Pardo, Adelya Urmanche, Hannah Gash, Jaclyn Wiener, Nicholas Mason, Sherilyn Wilman, Keagan Francis, and Alexa Decker

**Montclair State University
Department of Psychology
Speech Communication Laboratory
pardoj@montclair.edu**

Description

The Montclair Map Task Corpus of conversations in English comprises audio recordings and associated transcription annotation files for 48 pairs of native English speakers who engaged in a conversational task in the Speech Communication Laboratory at Montclair State University in New Jersey. This corpus was collected in March-May of 2013 (first 24 pairs) and in May of 2014 (last 24 pairs) as part of a large-scale study of phonetic convergence in spoken communication funded by the National Science Foundation PAC program (BCS #1229033, Jennifer Pardo, Principal Investigator). The task involved paired iconic maps with labelled landmarks, and the goal of the task was for the talkers to reconcile differences across the maps in the composition and location of the landmarks. Thus, recordings comprise natural conversational speech in which newly acquainted talkers discuss and repeat pre-determined landmark label phrases across 6 pairs of maps. Conversations ranged from 16 to 62 minutes in duration, averaging 30 minutes.

The corpus is useful for anyone interested in investigating naturally-produced, unscripted task-oriented conversational speech. The segmentation of speaking turns, pauses, fillers, backchannels, and overlapping speech events permits assessment of conversational turn-taking dynamics. The quality of the audio recordings, with 2-channel talker separation in a sound-proof booth, permits detailed analysis of acoustic-phonetic attributes. In particular, the set of 79 landmark label phrases was designed to elicit a full set of English vowel sounds as well as to provide opportunities to observe various phonological phenomena such as regressive place assimilation (e.g., green bay) and a variety of dialect markers (e.g., greasy wash water). The landmark label list and task maps are included as separate .pdf files.

To summarize, the corpus comprises 48 .wav files containing 2-channel audio recordings, 48 .TextGrid files with segmentation and annotation data for each audio recording (viewable with audio files in Praat), and 48 .csv files containing data extracted from the .TextGrid annotations. Additional supporting documents include a .pdf file with 12 map task images (6 pairs), a .pdf file containing a listing of all landmark label phrases, a .csv file with demographic information for all talkers, and a reprint of a journal article detailing corpus collection methods and preliminary analyses.

Recording Details

The talkers were all native English speakers (some were bilingual) who had resided in New Jersey for at least three years prior to participating. They provided informed consent following an IRB-approved protocol and specifically agreed to sharing their recordings in other studies. All talkers were part of the Montclair State University community and were paid \$20 for completing the recording session. The talkers were randomly assigned to 16 same-sex female, 16 same-sex male, and 16 mixed-sex pairings. Talkers completed the map matching task with 6 pairs of maps at their own pace, and they could not see each other or their partner's maps. Demographic information for each talker is included in a .csv file.

Each audio file comprises a 2-channel recording of a single pair of talkers. Each pair of talkers was recorded in the Speech Communication Laboratory at Montclair State University in an Acoustic Systems sound-proof booth using head mounted AKG microphones connected to a Macintosh computer. Recordings were made using SoundStudio software with 16-bit, 44.1 kHz sampling rate and saved in .wav format.

Transcription Protocol

Trained research assistants segmented and orthographically transcribed each audio file following a modification of the QRTR Broadcast English (XTrans Format) conventions provided by the Linguistics Data Consortium. The associated segmentation and transcription parameters are saved in Praat .TextGrid files that annotate the .wav files when opened together in Praat software. Each .TextGrid file comprises 4 tiers which specify 1) transcription content, 2) content type, 3) talker identification, and 4) map number. In addition, segmentation and transcription data from the .TextGrid files has been extracted into individual .csv files.

Segmentation of audio files delineated speaking turns (utterance by a single talker accomplishing a communicative goal), inter-turn intervals (time from the end of one talker's turn to the beginning of their partner's turn), inter-map intervals (time from the end of discussing one map pair to the beginning of discussing another map pair), pauses (within-turn silent intervals > 100 ms), fillers (pause fillers such as uh, um, etc.), backchannels (utterance produced by a partner during the main talker's turn that does not initiate a new turn, e.g., uh-huh, okay, etc.), overlapping speech (events in which both talkers speak simultaneously), non-speech utterances, and landmark label phrases. Note that only landmark label phrases are segmented, not individual words or phonemes. All speech within a turn that is not a landmark label phrase is transcribed in a single segment labelled *context* (if there were no pauses or other events). For example, if a talker said "I don't have a west lake," the utterance would be segmented into two events, "I don't have a" and "west lake." The first segment would be transcribed verbatim in the content tier and labeled *context* in the content type tier, and the second segment would be transcribed in the content tier and labeled *landmark* in the content type tier.