

# An international English speech corpus for longitudinal study of accent development

Rosemary Orr<sup>1</sup>, Hugo Quené<sup>2</sup>, Roeland van Beek<sup>1</sup>,  
Thari Diefenbach<sup>1</sup>, David van Leeuwen<sup>3</sup>, Marijn Huijbregts<sup>3</sup>

<sup>1</sup>University College Utrecht, Utrecht University, The Netherlands

<sup>2</sup>Utrecht Institute of Linguistics OTS, Utrecht University, The Netherlands

<sup>3</sup>Department of Language and Speech, Radboud University, Nijmegen, The Netherlands

r.orr@uu.nl, h.quene@uu.nl

## Abstract

If English is used intensively as a lingua franca in a multi-language community, do speakers then converge towards a single common accent? This speech corpus allows for longitudinal study to investigate the question of convergence by means of repeated speech recordings of students at an English-language college over a period of 5 years. We describe the content and collection of the corpus and the type of research that is envisaged, as well as tools used to manage and analyze the recordings, including automatic phone recognition for prosodic analyses; and intelligibility experiments using the SRT method.

## 1. Introduction

When people from native and non-native backgrounds come together, and all speakers use English as a lingua franca, then how do their English accents change over time?

Do native speakers drift away from their native pronunciation standards? Do non-native speakers become more native-like, and does interference from their L1 decrease over time? Is the speaker's accent of English related to their intelligibility and subjective accentedness? The English-language international Liberal Arts and Sciences College, University College Utrecht (UCU) in the Netherlands, provides an interesting environment to investigate such questions, as we explain below.

This paper provides some background of a longitudinal study about the development of English accents, and it describes the speakers and data collection procedure of that study. In addition, we report on current analyses and preliminary results from the first round of recordings, looking into the use of tools from speech technology research. These tools are intended for automatic and semi-automatic processing of our recordings, in order to facilitate subsequent linguistic-phonetic research. At this point, the first cohort of 72 students has been recorded, and a few tools for analysis have been developed. A further group of 23 native English speakers has been recorded, for use as a control group, and as a listener group for intelligibility tests.

A core hypothesis in this project is that the native and non-native accents of UCU students will gradually converge to a single common international variety of English, which we call *UCU English accent*. This has implications, both social and linguistic, for the speech of this student group, and these will be investigated over the coming five years.

When talkers from different languages or dialects converse with each other, it seems that their dialects and accents tend to converge. This phenomenon has been observed for dialects

of British English [6] as well as for dialects of Dutch in the IJsselmeerpolders [13]. This convergence, and its opposite, namely divergence, have been described by the Communication Adaptation Theory [9]. According to this theory, younger talkers are more susceptible to this outside pressure on their dialect or accent than older talkers are. Hence, university students provide an excellent group for the investigation of this phenomenon. Previous research involving university students has focused on native-language speakers of *Northern* and *Southern* varieties of British English [6], and we extend this approach, to include both native (L1) and non-native speakers (L2) of English from the international body of students at University College Utrecht. To this end, we have recorded the speech of new students in their first weeks of study at the College, and we will repeat these recordings at four subsequent moments during the students' three-year stay at UCU. To our knowledge, this study is innovative because of the unique language environment at UCU (see below), as well as the large-scale phonetic analyses, speech perception experiments, and ASR as an analysis tool.

We expect that the factors affecting the emergence of a UCU accent of English will include the sort of English spoken by teachers in the classroom setting as well as the social groups formed by the students. Students tend to be very involved in the various campus committees within the Student Association, and their social groups are often formed around these. These observations lend themselves to sociolinguistic research, where the influence of the linguistic environment of the social and academic groups on the emergent accent can be explored. In particular, since the cultural and linguistic profiles of the social groups on campus change with each year, we might expect the UCU accent to be slightly different for each three-year cohort. This is particularly the case in social groups where a Dutch L1 is not prevalent.

Further opportunities for sociolinguistic research arise in the exploration of attitudes to the development of an accent of English. For example, it has been shown by [8] that hyperaccommodation to prestige forms of English may evoke negative reactions from listeners, both native and non-native. Most students have a strong desire to achieve a native-sounding accent [14, 11]. It is conceivable that some students will have a prestige accent as their target, while others will not. Listener attitudes to speaker accents, coupled with listener appraisal of nativeness of speaker accents may shed light on the type and degree of accommodation present at UCU, as well as the reactions that such accommodation provokes in listeners from within and outside the campus community.

In terms of models of speech perception and language learning, the corpus will be used to examine in how far the *Critical Period Hypothesis (CPH)* can be applied to our speaker group. We will also look at Flege’s *Speech Learning Model*, comparing languages of similar and dissimilar prosodic, phonological and phonetic composition, looking for evidence of phonetic category assimilation and dissimilation in this international environment.

## 2. The corpus

### 2.1. Speakers

#### 2.1.1. Groups for longitudinal study

The majority of the speakers are students at UCU, an English-language college in a Dutch-speaking host environment in Utrecht, the Netherlands. The students (ages approx. 17–24 years) originate from all over the world. The student demographics are quite stable, with 60% students from the host country and 40% from elsewhere. Students are required to speak English in all their interaction with tutors and teachers. Students live on campus during their entire three-year degree program, and in culturally and linguistically mixed social settings, English is also the language of choice.

The majority of the students does not speak English as a native language. In a crude approximation, one might say that the overall L2 English accent is influenced by British and US accents of English, depending on the student’s teacher at high school, on the student’s experiences abroad, and on the degree of exposure to US and UK popular culture exported through books, movies, popular music, etc. Further, the relatively large group of native Dutch speakers of L2 English in the student population of UCU is likely to have a relatively strong influence on the UCU English accent. Other native speaker groups are smaller in size, and are therefore predicted to exert a relatively small influence.

At UCU, then, English is the only lingua franca, and UCU students and teachers interact intensively in English. Students use English all day, and many students use the surrounding language (Dutch) only seldom or never at all. The student group contains few native speakers of English, many native Dutch speakers of L2 English, and a large variety of other L1 speakers of L2 English. Outside-college factors on accent development can be largely excluded.

#### 2.1.2. Native English speaker group

In addition to the incoming first semester students, 23 native speakers of English were recorded in January 2011. These speakers produced similar speech data to that of the international student speakers, to be used as control data, as well as in intelligibility testing. The native speakers, therefore, record extra sentence sets for intelligibility testing. These speakers were taken from the group of exchange students, who visit UCU for a semester abroad. Most exchange students come from the United States, although there is a smaller group of native English speaking exchange students from the U.K. and South Africa. Over the coming years, such exchange students will be recorded and added to the initial native speaker group. We aim for around 125 native speakers by the end of the five-year recording period. The current language profile of the native speaker group is such that some speakers appear to have some experience already with a Dutch-speaking environment. For this reason, we intend to split this group in two at a later stage, one of which will be free of the influence of the Dutch language environment at the time of recording.

Table 1: Five-year schedule of speaker recordings for longitudinal study, showing current and estimated numbers of speakers

Date	Cohort I	Cohort II	Cohort III	Total
Sept 2010	72	-	-	72
May 2011	60	-	-	60
Sept 2011	60	70	-	130
May 2012	55	60	-	115
Sept 2012	-	60	70	130
May 2013	50	55	60	165
Sept 2013	-	-	60	60
May 2014	-	50	55	105
Sept 2014	-	-	-	-
May 2015	-	-	50	50
Total				887

### 2.2. Speech recordings

#### 2.2.1. Groups for longitudinal study

Speakers are recruited from the incoming first-semester students in September 2010, 2011 and 2012. Within six weeks of their arrival at UCU, speech from the participating students is recorded during an hour-long session. Speakers are recorded again at four further moments during their three-year period of undergraduate study, namely at the end of semester 2, before summer break; at the beginning of semester 3, after summer break, spent off campus; at the end of semester 4; and finally at the end of semester 6, just before graduation. Table 1 shows the recording schedule with the intended number of speakers.

Each session consists of seven parts, recorded in a single sound file. These seven parts include two read texts (part of *the Rainbow passage* [7] and one of the well-known *Æsop’s fable The boy who cried wolf*); five sentences for prosodic analysis; three sets of sentences for intelligibility testing; prepared but not practiced speech on two topics; and free conversational speech. If English is not the speaker’s L1, they are asked for a sample of prepared speech in their native language as well.

The recording sessions take place in a quiet furnished office room where the subject is surrounded by sound-absorbing screens to reduce room echo. All electronic and electrical equipment is turned off during the recordings, except, of course, for the recording equipment itself.

#### 2.2.2. Native English speaker group

The same recording procedure is followed for this group, except that they are required to record 10 sets of 13 sentences for intelligibility testing, instead of just 3.

### 2.3. Recording equipment

Speech recordings are made via a Saffire Pro 40 multichannel AD converter and preamplifier, using Audacity<sup>1</sup> software for recording and editing. Speech is recorded via 7 microphones (Sennheiser ME64/K6p) placed in different locations in the recording area, namely in front of the speaker, behind, left, front-left, right, front-right, above, and also via a close-talking microphone (Sennheiser Headset HSP 2ew). The data from multiple microphones are intended for a different unrelated project on automatic speaker recognition from longitudinal data. Figure 1 shows a schematic view of the setup.

<sup>1</sup>Audacity is open source software for recording and editing sounds, see <http://audacity.sourceforge.net/>



time-varying amplification factor for that sentence token. Intelligibility tests using adjusted target sentences are now carried out, with materials from L1 Dutch speakers of L2 English, and with listeners being native speakers of Dutch, English or some non-Germanic L1.

A preliminary comparison of L1 English listeners, listening to both L1 English and L1 Dutch speakers shows that, while both speaker sets are quite intelligible, there is degradation of intelligibility of the Dutch speakers. As the corpus expands and the test results for different L1 groups of both listeners and speakers become available, we will be able to measure changes in intelligibility - for better or for worse - as the students progress through their undergraduate studies.

### 3.4. Investigation of other automatic methods for analysis

We will also investigate whether changes in intelligibility and accentedness can be assessed automatically and objectively, by means of automatic speech recognition techniques. For some clinical areas of intelligibility assessment, such tools can produce efficient objective measures, see e.g. [12] for similar assessments of intelligibility in clinical settings. It is interesting to explore whether measures such as word error rates (WER) from speech recognition tests could be used to assess the improvement or disimprovement of intelligibility of non-native speech, and to correlate such measures with human listeners' evaluations. For this, we intend to train acoustic models on speakers from final-semester students as well as models conditioned on British and US English exchange students, who come from the UK or from the US to study at UCU for one semester, and who have not been exposed to the UCU English accents.

## 4. Other envisaged investigations

We plan to carry out sociolinguistic studies, following the students' social groups, and establishing a cultural and linguistic profile of college committees and associations. For example, there are fraternities and sororities that are almost exclusively Dutch, while some committees that focus on intercultural activities are international. In the core members of these two groups, we might expect different accents of English to emerge.

Still within a sociolinguistic framework, we will examine attitudes to strong, mild and hypercorrect accents. Young speakers often try hard to achieve native-like pronunciations in their peer groups. Since there is no native pronunciation of a UCU accent of English, we might expect that some students will actively aim for native-like pronunciations of General American or RP English. As [8] demonstrated, hyperaccommodation may not always be desirable, from the point of view of social acceptance, and hyperaccommodation to a non-peer group prestige accent may be even less attractive.

We will also look for elements of the emergent accent of English which do not typically belong to a standard English phonetic inventory. Given the predominance of Dutch as L1, expected candidates might be lack of or reduced aspiration of voiceless stops, a more retracted alveolar fricative version of /s/, or substitution of less central vowels for *schwa* in unstressed syllables (see, for example, [3]).

When the corpus is complete, the recordings will be transcribed. We intend to make the data available to the speech community at large. Other corpora with non-native English recordings can be used for comparison with ours. The longitudinal nature of the data may be interesting for use in the area of speaker characteristics and speaker recognition, and for the effect of speaker adaptation on speech recognition.

## 5. References

- [1] P. A. Barbosa, P. Antares, and L. S. Silveira. Unifying stress shift and secondary stress phenomena with a dynamical systems rhythm rule". In *Proceedings of the 2nd International Conference on Speech Prosody*, Nara, Japan, 2004.
- [2] Paul Boersma and David Weenink. Praat: doing phonetics by computer. version 5.x. retrieved from <http://www.praat.org/>, 2010.
- [3] B. Braun, K. Lemhofer, and Nivedita Mani. Perceiving unstressed vowels in foreign-accented english. *Journal of the Acoustical Society of America*, 129(1):376–387, 2011.
- [4] H. Quené. Andante of allegro? verschillen in spreektempo tussen vlamingen en nederlanders. *Onze Taal*, 77(6):179–181, 2008.
- [5] H. Quené and L. E. van Delft. Non-native durational patterns decrease speech intelligibility. *Speech Communication*, 52(11-12):911–918, 2010.
- [6] B. G. Evans and P. Iverson. Plasticity in vowel perception and production: a study of accent change in young adults. *Journal of the Acoustical Society of America*, 121(6):3814–3826, 2007.
- [7] G. Fairbanks. *Voice and Articulation Drillbook*, chapter 5, page 127. Harper and Row, 1960.
- [8] Peter Garrett. Accommodation and hyperaccommodation in foreign language learners: contrasting responses to french and spanish english speakers by native and non-native recipients. *Language and Communication*, 12(3/4):295–315, 1992.
- [9] H. Giles, N. Coupland, and J. Coupland. Accommodation theory: communication, context and consequence. In J. C. H. Giles and N. Coupland, editors, *Contexts of Accommodation: Development in Applied Sociolinguistics*, pages 1–68. Cambridge University Press, 1991.
- [10] E. Grabe and E.L.Low. Durational variability in speech and the rhythm class hypothesis. In C. Gussenhoven and N. Warner, editors, *Papers in Laboratory Phonology VII*, pages 515–546, Berlin, 2002. Mouton de Gruyter.
- [11] J. Jenkins. *English as a Lingua Franca: attitude and identity*. Oxford University Press, Oxford, 2007.
- [12] A. Maier, T. Haderlein, U.Eysholdt, F. Rosanowski, A. Batliner, M. Schuster, and E. Nöth. Peaks - a system for the automatic evaluation of voice and speech disorders. *Speech Communication*, 51(5):425–437, 2009.
- [13] H. Scholtmeijer. *Het Nederlands van de IJsselmeerpolders*. Kampen Mondiss, 1992.
- [14] I. Timmis. Native-speaker norms and international english: a classroom view. *ELT Journal*, 56(3):240–249, 2002.
- [15] R. van den Doel. *How friendly are the natives? An evaluation of native-speaker judgements of foreign accented British and American English*. PhD thesis, Utrecht Institute of Linguistics OTS, Utrecht University, Utrecht, The Netherlands, 2006.
- [16] S. J. van Wijngaarden. Speech intelligibility of native and non-native speech. In H. J. M. Steeneken, D. A. van Leeuwen, and S. J. van Wijngaarden, editors, *Multi-lingual Interoperability in Speech Technology*. ISCA, 1999.
- [17] L. White and S. Mattys. Calibrating rhythm: first language and second language studies. *Journal of Phonetics*, 35(4):501–522, 2007.