

## CHAPTER 15

# D-LUCEA: Curation of the UCU Accent Project Data

Rosemary Orr<sup>†</sup> and Hugo Quené

### ABSTRACT

The UCU Accent Project was set up in 2010 to collect a wide variety of non-native and native accents of English in an environment where English is the *lingua franca*, namely an international liberal arts and sciences college in Utrecht in the Netherlands. The recordings were made longitudinally over the three years of undergraduate study, and four cohorts of students were recorded in total. This yielded over 1,000 speech recordings over a six-year period in which the development of both native and non-native English accents in a non-native environment can be examined. In order to facilitate sharing the data with the wider research community, the D-LUCEA project undertook to curate the data. For each recording, the relevant concomitant metadata was produced, giving information to users of the database about the speaker, the technical specifications, the kinds of speech material recorded, and so forth. The project was funded by CLARIN, and specific CLARIN tools for curation were made available to us, including the Component Metadata Infrastructure (CMDI). To date, all of the speech data has been processed such that the metadata is available, and research is already running on this corpus, on topics as varied as prosodic convergence, L1 phonetic drift and phone convergence. Further plans include work with speaker recognition, accent recognition and models of language learning such as Flege's Speech Learning Model, the Critical Theory Hypothesis, and the Perceptual Assimilation Model.

### 15.1 Introduction

This chapter describes the UCU Accent project, and the curation of the resulting D-LUCEA database, making the recorded speech data with its concomitant metadata widely available to the

---

<sup>†</sup> Deceased.

---

#### How to cite this book chapter:

Orr, R and Quené, H. 2017. D-LUCEA: Curation of the UCU Accent Project Data. In: Odijk, J and van Hessen, A. (eds.) *CLARIN in the Low Countries*, Pp. 181–193. London: Ubiquity Press. DOI: <https://doi.org/10.5334/bbi.15>. License: CC-BY 4.0

research community at large. The data-curation project was funded by the Dutch partners in the pan-European CLARIN project, whose goals are to facilitate precisely this kind of work.

We describe some of the research that has been made possible via this project, as well as current plans for employing a similar method for the curation of data in a new speech accent corpus, Sprekend Nederland.

### 15.1.1 *The UCU Accent Project*

Evidence from research over the last few decades indicates that when talkers from different language or dialect backgrounds converse with each other, their dialects and accents tend to converge. This phenomenon has been observed for dialects of British English (Evans and Iverson 2007) as well as for dialects of Dutch in the IJsselmeerpolders (Scholtmeijer 1992) and has been observed in phonology, phonetics and stylistics (Pardo 2006).

Such convergence, and its opposite, namely divergence, are described by the Communication Adaptation Theory (Giles et al. 1991). According to this theory, younger talkers are more susceptible to this outside social 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 speakers of Northern and Southern varieties of British English (Evans and Iverson 2007). However, while social context can be important, convergence has also been observed without social context in word shadowing tasks (Goldinger 1998) and Trudgill (2004) suggests that, in line with the general human tendency to act like one's social peers, accommodation can be subconscious and automatic as well as conscious.

It is interesting to consider what happens in this respect when the common language is not a native language for the majority of speakers. When people from native and non-native backgrounds come together, and all speakers use, for example, 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? Does increasing proficiency in the L2 cause attrition in the L1? Is the speaker's English accent related to their intelligibility and subjective accentedness? And how stable are the speaker characteristics across L1 and L2?

The international *University College Utrecht* (UCU) in the Netherlands provides an ideal environment to investigate these kinds of adaptation, being an international body of students that includes both native (L1) and non-native speakers (L2) of English. To explore these questions, we set about collecting speech from students at the college at different moments during their three-year undergraduate program, covering four consecutive cohorts over a period of six years. Along with the speech data, we have recorded a rich set of metadata, including technical data about the equipment, speaker and facilitator data, session data, and logbook observations about each recording.

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 the UCU English accent. The convergence of a group of non-native accents to an international non-native variety has implications, both social and linguistic, for the speech of this student group, and is the overarching theme in our work on this project.

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 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 Garrett (1992) that hyper-accommodation 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 (Timmis 2002; Jenkins 2007). 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 accuracy of speaker accents may shed light on the type and degree of accommodation present at UCU, as well as the affective responses and intelligibility scores resulting from such accommodation in listeners from within and outside the campus community.

## 15.2 D-LUCEA: Sharing the Data in the Research Community

The possibilities for research on this speech data likely extend to a great many areas, including sociolinguistics, sociology, phonetics and phonology, and speech technology.

It has been our intention from the start to make the data freely available for scientific research, so that colleagues from anywhere can use our data to verify our findings or to explore different aspects or themes themselves. This raises the question of how to curate and distribute the speech data and metadata in a way that makes it maximally useful to the broad range of users that we envisage.

In order to be maximally useful, the format of the data and metadata files must facilitate interoperability across different kinds of technological infrastructures and collaboration across different research disciplines. The format should be robust against developments in and variations of software and hardware, and should meet an international standard.

In general terms, the curation of the data consisted of creating a general metadata profile to describe a generic speech recording, and then for each actual speech recording, creating a specific instance of that profile and linking it to the speech recording in question. This information, including the speech data, was then made available for download to the research community at large. The corpus was given the name *D-LUCEA*, for *Database of the Longitudinal Utrecht Collection of English Accents*.

In order to describe the curation process, we first give a description of the procedure for the data collection procedure, and then we describe the process of organising the metadata, and linking the speech recordings to that metadata in order to make it available for general access.

### 15.2.1 *Recording Procedure, Speech Data and Concomitant Metadata*

#### 15.2.1.1 *Recording Setup*

Recording sessions took place in a quiet furnished office, with one or more facilitators and a speaker participant. Recordings were made on eight different channels. Figure 15.1 shows a schematic view of the setup where the positioning of each microphone is clearly marked. Microphone 1 is a close-talking headset microphone.

For each recording session, then, eight speech files were produced. The metadata associated with each recording is specific to the particular microphone channel.

#### 15.2.1.2 *Timing of the Recordings*

Between August 2011 and June 2016, four cohorts of students took part in the project. For each cohort, between 60 and 80 students took part in at least the first recording. Recordings were made at five moments, or rounds during the students' period of study, namely at the beginning and end of

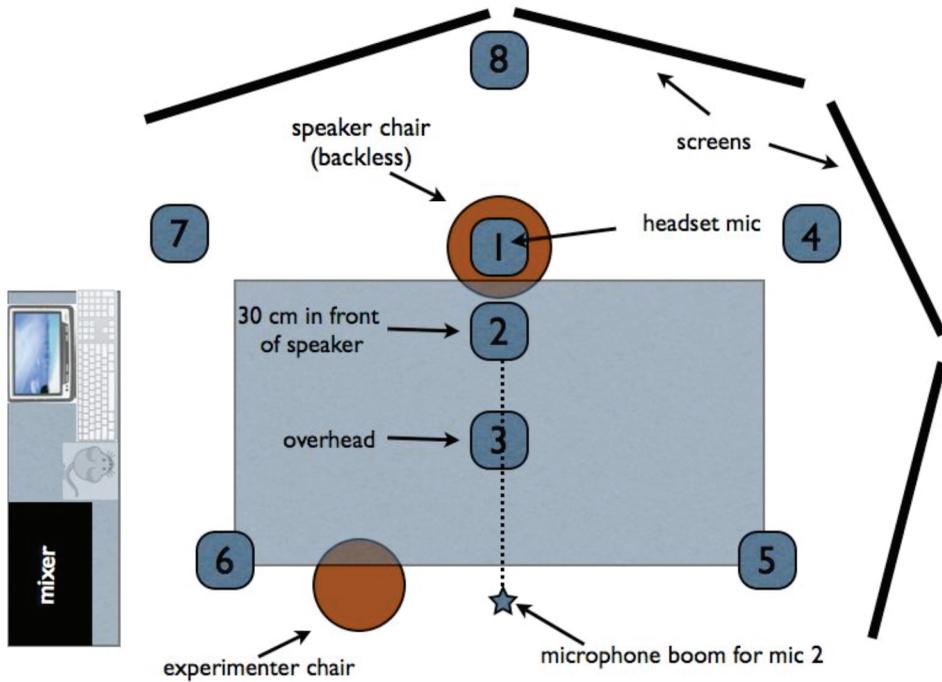


Figure 15.1: Schematic view of the recording setup.

Date	Cohort I	Cohort II	Cohort III	Cohort IV	Total
Sept 2010	79	-	-	-	75
May 2011	67	-	-	-	67
Sept 2011	60	78	-	-	138
May 2012	35	66	-	-	101
Sept 2012	-	61	72	-	122
May 2013	50	47	55	-	152
Sept 2013	-	-	51	58	107
May 2014	-	53	34	37	124
Sept 2014	-	-	-	40	40
May 2015	-	-	47	36	83
Sept 2015	-	-	-	-	0
May 2016	-	-	-	39	39
Total					1,048

Table 15.1: Six-year schedule of speaker recordings for longitudinal study, showing the number of speakers who participated in each round.

each college year, with the exception of the beginning of the third year, or fifth semester. Table 15.1 shows the recording schedule over the six years of the collection of the corpus.

### 15.2.1.3 Session Information

For each session, the following information is provided;

- recording ID, incorporating subject number and round number
- recording round, being one of round 1 to 5 for the speaker

- channel number, equivalent to the microphone number in the setup scheme above
- recording date
- whether an audible separator was used, and if so, what kind<sup>1</sup>.

#### 15.2.1.4 Speaker Task Information

The speakers were required to perform between 9 and 12 speaking tasks in each round, as outlined in Table 15.2 below. Some explanatory notes are also given for particular tasks, where relevant.

Most of the tasks were present from the very first recording, but others were added in order to produce data for comparison with other accented-speech corpora, in particular, the OSCAAR corpus and the ALLSTAR corpus.<sup>2</sup> Specifically, the initial design did not include the articles from the Universal Declaration of Human Rights; these were introduced at the second round of recordings of the first cohort.

Task 1 was deleted from the recordings before publishing in order to preserve privacy. It allowed for a double-check on the speaker information per recording.

Task 4, *The Boy who Cried Wolf*, was initially a long passage with few shibboleths. From the second round of recordings from Cohort I, a second version, shorter and containing shibboleths, replaced the original one. Both versions can be found in Appendix IV, where the texts for tasks 2 to 7 are provided.

The substitution of a text with shibboleths was intended to elicit the different substitutions used by L2 speakers of English, and to examine whether and how these change over time.

Task 5 refers to sentences from Van Wijngaarden et al. (2002) for quantifying intelligibility of speech in noise for non-native listeners. There are 10 sets of 13 sentences, also to be found in Appendix IV. Native speakers of English were generally asked to speak all 10 sets. Non-native speakers were asked to speak between 3 and 4 of these sets. This was done in order to make sure

Task	Task Description
1	Speakers state their name, date and time
2	Short extract from the Rainbow Passage (Fairbanks 1960)
3	'Please Call Stella' (Weinberger 2013)
4	'The Boy who Cried <i>Wolf</i> ' (Deterding 2006)
5	Balanced sentence sets for intelligibility testing (van Wijngaarden 1999)
6	Five sentences for investigating rhythm (White & Mattys 2007)
7	Extract: Declaration of Human Rights in L1 (Universal Declaration of Human Rights 2013)
8	Extract: Declaration of Human Rights in English (Universal Declaration of Human Rights 2013)
9	2 minute monologue L1, informal free topic
10	2 minute monologue English, informal free topic
11	2 minute monologue English, formal free topic
12	3 minute dialogue English with the facilitator, free topic

**Table 15.2:** The speaker tasks that could be required in a single recording session.

<sup>1</sup> After the first round of recordings, an audible separator was introduced between the tasks. The choice for an audible separator stems from the nature of the recording setup. It is not only a signal for separating tasks, but it is also a prompt for the participant, making clear when they should speak. It varied between the sound of a tap on a glass, a high-pitched recurring ping, or a bell.

<sup>2</sup> OSCAAR corpus: <https://oscaar.ci.northwestern.edu/overview.php>  
ALLSTAR corpus: [http://groups.linguistics.northwestern.edu/speech\\_comm\\_group/allstar/](http://groups.linguistics.northwestern.edu/speech_comm_group/allstar/)

that not all non-native participants had read all texts. In this way, the participants could also take part in tests to assess the intelligibility of other speakers in the project. Cohorts I and IV spoke sets 1 to 3; Cohort II spoke sets 4 to 6; Cohort III spoke sets 7 to 10.

The metadata information per recording gives the tasks that were spoken for that recording, the order in which they were spoken, as well as the approximate start and end times for each task. Information included per task is as follows:

- modality (spoken)
- interactivity (whether interactive, semi- or non-interactive)
- whether spontaneous, semi-spontaneous or planned
- whether elicited or spontaneously generated
- whether monologue or dialogue

#### 15.2.1.5 *Speakers in the Recordings*

The people recorded speaking in the project include the speaker who is producing the speech tasks, as well as the facilitators, who play a role not only in guiding the speaker through the tasks, but also in engaging in dialogue with the speaker during each session.

The speakers are mostly students, plus a few staff or faculty members at University College Utrecht UCU). The facilitators are faculty, staff and graduate or undergraduate students at UCU or at Utrecht University (UU).

#### 15.2.1.6 *Speaker Characteristics*

For each speaker, a number of aspects of their exposure to different languages, physical characteristics, musicality, hearing ability and language practice are considered relevant to many of the questions that we envisage as applying to this dataset.

The lists below indicate the metadata related to the speaker. Most of the information regarding language usage was gathered from a questionnaire that the students filled in on entry to the project. A second questionnaire was filled in on completion of their degree with new questions which captured information that arose during the three-year period, for example, the student's major or possible minor, or where and when they had gone abroad for a semester.

The questions related to language learning from the first questionnaire were repeated in the exit questionnaire, and if there was a difference in the answers, the second answer was taken as the representative one. The reason for this is that students at UCU are required to learn another language to a good level of proficiency, and after three years they may have become less proficient – either comparatively or actually – in other languages that they spoke on entry to the college.

One of the interesting issues in assessing language exposure and proficiency was that of defining the native language or languages. This particular student population contains members for whom it is difficult to define a native language. Languages which were learned first were not always the dominant languages, and were sometimes either forgotten or underdeveloped.

For example, one participant has a father speaking one language X, a mother speaking language Y. Her father's language X was the first language she learned, albeit poorly, and she could understand but not speak her mother's language Y. She regarded neither X nor Y but English as her native language, although she only learned English via an English-speaking Russian nanny and an English-language day care centre in Beijing.

This was not the only such case, and because of this difficulty in establishing a native language, we opted to ask about languages learned before the speaker was eight years old. The choice of this age is fairly arbitrary, but does allow for childhood development of fluency in a language.

For each speaker, the following general information is available:

- personal information: sex and date of birth
- physiological information: height and weight
- audiometric information: for both ears, the hearing threshold for frequencies between 250 Hz and 8 kHz
- a self-assessment of musical, language and hearing abilities
- languages learned before eight years of age
- all languages spoken by the speaker
- situations in which each language is spoken
- English language information regarding age of learning, years of experience and proficiency

Where the speaker is a student, information is provided on their curriculum. This includes:

- major(s) or main field(s) of study
- minor(s)
- academic disciplines
- whether the student went on an exchange semester abroad, and if so, where, and what language was spoken there
- date of entry to the college
- graduation date

Students were also asked to undergo an audiometric test at the end of their final recording. Hearing threshold values (in dB) were measured for key frequencies<sup>3</sup> for both the left and right ears.

#### *15.2.1.7 Facilitator Characteristics*

A recording session could be attended by more than one facilitator. Initially, the facilitators worked in pairs to establish and monitor a standard protocol for recordings.<sup>4</sup> As the project progressed, new facilitators joined and for purposes of monitoring and instructing, a second, more experienced facilitator was present. Many sessions, however, were facilitated by just one person.

Similarly to the speakers, facilitator information includes the general information above, along with the following:

- name
- affiliation (UCU or UU)
- whether they were the primary facilitator
- whether they were a student

#### *15.2.1.8 Sound File Information*

Information on the sound file itself is provided as follows:

- creation date
- quality of recording – the close-talking headset microphone number 1 was of very high quality, indicated as 2 on a scale of 1 to 7 (where 1 indicates highest quality); the remaining microphones were also of very good quality, but we rated them as having a quality of 3 on this scale

---

<sup>3</sup> at 0.25, 0.50, 1, 2, 3, 4, 6 and 8 kHz.

<sup>4</sup> This protocol is included in the metadata as an accompanying file.

- recording condition (quiet furnished office)
- recording platform software
- recording platform hardware
- size of the file
- sampling frequency (44.1kHz)
- byte order (little endian)
- bit resolution (32)
- speech coding (pcm)
- mime-type (audio/wav)

#### 15.2.1.9 Metadata Structure

A unique feature of the current project is that the data are longitudinal by design. One primary talker or informant is recorded in at least one and at most five rounds, with various interviewers across sessions, and with a variable number of interviewers being present at a single session. All this information is relevant and needs to be accessible. Hence a simple structure of *one session: one talker: one interviewer* does not suffice. Moreover, information may change between sessions: a talker can mention Russian as a native language in session 1, but he or she may no longer mention Russian as a native language in session 5, three years later.

The metadata described above were obtained from various sources. Immediately after the first recording (in the same session), the entry questionnaire was administered. Notes were logged during each recording about any special circumstances and about topics during the monologues. Immediately after the last recording, hearing was measured (the dB threshold values were stored in a spreadsheet) and the exit questionnaire was administered. Technical details of the audio files were also stored in a separate file.

In creating our metadata scheme, we made maximum use of CLARIN's Component Metadata Infrastructure (CMDI). Most of the metadata categories were already in existence, and where necessary, within the CMDI structure, we created new ones.<sup>5</sup>

#### 15.2.2 Linking Speech Recordings and Metadata

The various metadata were combined from all these metadata sources into a single annotation profile named *lucea.xsd*. A custom-built Python script extracted relevant metadata from various sources, checked these metadata for consistency with ISOCat and for internal consistency, and wrote the metadata into an XML file, compatible with the CMDI metadata scheme.<sup>6</sup> These XML files constitute a hierarchy (tree), with multiple metadata files that correspond to multiple audio files in a session, and with multiple sessions nested under a speaker. This meant that relevant information had to be copied to subordinate nodes of this branching hierarchy. Finally, the Python script inserted in each XML metadata file a persistent resource link to the appropriate audio file.

### 15.3 Current Research Using this Corpus

The collection of the LUCEA speech corpus was completed in May 2016. Some preliminary research has been conducted to explore the potential that the corpus has for answering the questions with which we started out. This initial exploratory research has yielded some interesting results, and the plans for comprehensive work on these questions are taking shape.

<sup>5</sup> A list of all 30 new concepts and properties is available from the ISOCat repository of data categories, at <http://www.isocat.org/rest/dcs/649>.

<sup>6</sup> <https://www.clarin.eu/content/component-metadata>

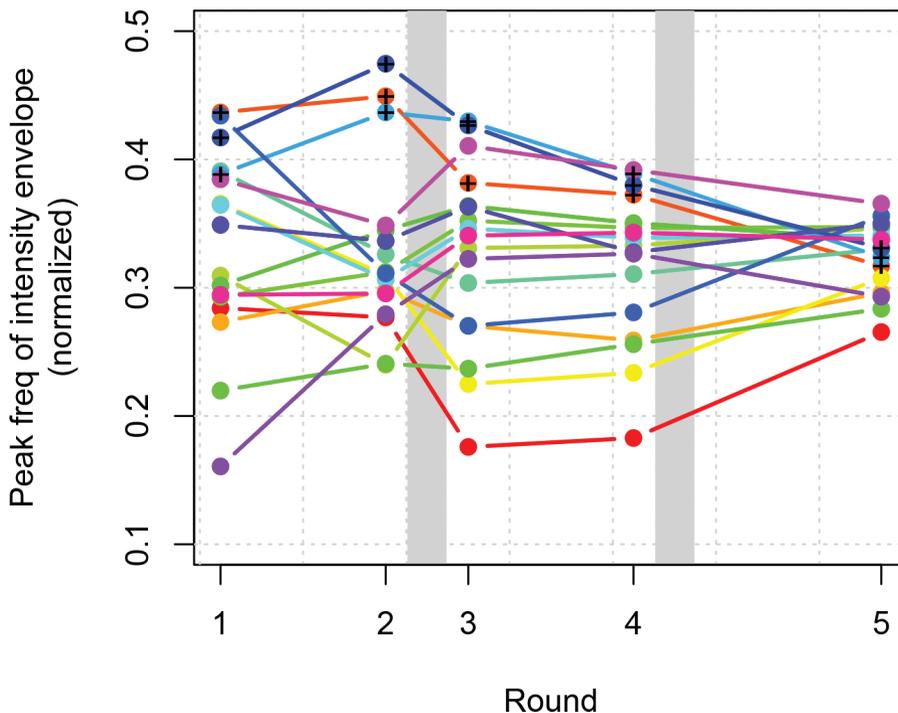
### 15.3.1 Prosodic Convergence

For example, in looking at prosodic patterns over time, Quené and Orr (2014) found that at least one aspect of the speakers' prosodic behaviour seems to converge over time. In this study, the normalised peak frequencies in the spectrum of the intensity envelope were compared for five English sentences, taken from and studied by White & Mattys (2007). Eighteen speakers from the corpus were studied, of whom fifteen talkers declared themselves as native speakers of Dutch, and one talker each as a native speaker of Russian, Vietnamese, and German. Three speakers (one female, two male) also regarded themselves as L1 English speakers, that is, as bilingual Dutch and English.

In Figure 15.2, the peak frequencies in the final recording for these subjects can be seen to converge. In a linear mixed-effects regression analysis of this data, English L1 speakers initially showed significantly higher peak frequencies in the intensity envelope than the Dutch L1 speakers. We interpret this as reflecting the stronger reduction of unstressed syllables in English as compared to Dutch. Over time, the values for the English L1 speakers move towards values in the centre of the range of converged values.

### 15.3.2 Intelligibility Across Time

The initial investigation of prosodic behaviour supports the results from other research, namely that speakers tend to accommodate to each other while talking. We might expect, then, that



**Figure 15.2:** Estimates of normalised peak frequencies in the spectrum of intensity envelope, broken down by round of recording (along abscissa, on approximate time scale) and by talker (with plussed symbols representing L1 English speakers). Shaded areas represent 2-month summer breaks during which talkers do not live on the UCU campus. Note that there is a full year gap between round 4 and round 5.

intelligibility – certainly within our student population – increases over time. We would predict that the intelligibility of post-accommodated speech is higher than that of pre-accommodated speech.

In a study of 45 speakers from the corpus, we measured the intelligibility over the first three rounds of recordings, that is, at the beginning of the first semester, the end of the second semester and the beginning of the third semester. The *subjects* were 9 English L1 speakers, 15 Dutch L1 speakers and 6 German L1 speakers. Of the *listeners*, 33 were Dutch L1 speakers, 5 were English L1 speakers and 7 were bilingual in Dutch and English. Intelligibility was measured using the Speech Reception Threshold measure, modeled on work by Van Wijngaarden et al. (2002), and using the sets of 13 test sentences.

The results showed that our subjects were indeed more intelligible in the second round of recordings than in the first round. However, this effect disappeared in the third round, which we attribute to the long two-month break away from the college community during the summer period after the first year. During this period, we suspect that speakers revert to their original ways of talking.

Notably, intelligibility after the third round was measured as poorer than after the first round. This has yet to be investigated, but it is possible that the first-round measurements, which were taken after an intensive introduction week in which incoming first semester students spend every day in activities with more senior students, already showed a small level of convergence.

### 15.3.3 L1 Phonetic Drift

The results of the initial prosodic investigation support the idea that not only might L2 English speakers' accents change over time, but so might also those of the L1 English speakers. Similarly, it may be that Dutch as L1 exhibits signs of phonetic drift, as a result of immersion in English over a three-year period. For this study (Orr et al. 2015), we looked at possible phonetic drift in word-initial /d/ and /t/, and the sibilant /s/, which are realised with audible phonetic differences in Dutch and English.

In non-clustered word-initial position, typical VOT values for Dutch voiceless stop /t/ and the English voiced stop /d/ are quite similar. Dutch voiced stops have a shorter lag time than their English counterparts, and English voiceless stops have a much longer lag time than their Dutch counterparts, being generally aspirated. Dutch has only one sibilant /s/ whereas English has two, namely /s/ and /ʃ/ (Boersma & Hamann 2008, Collins & Mees 2003). The articulation of the Dutch /s/ is described as being somewhere between the two English sibilants, having a more retracted position of articulation, a flatter tongue, and more lip rounding than the English /s/ (Collins & Mees 2003).

Because these particular phonemes exhibit phonetic, rather than phonemic differences in Dutch and English, it is interesting to explore them in the context of the Speech Learning Model (SLM; Yeni-Komshian et al. 2000), which suggests that the ability to perceive within-phoneme differences between an L1 and an L2 may drive the formation of a new phonetic category within a single phoneme. Conversely, if a speaker does not perceive the difference, this new category may not be formed at all, but both L1 and L2 values will assimilate towards each other.

We compared the two-minute L1 (Dutch) and L2 (English) monologues for 50 Dutch L1 speakers from the first two cohorts. We isolated all instances of /d/, /t/ and /s/ from the first and final rounds of recordings. VOT was measured as the period from stop burst to the onset of voicing, using manual segmentation in Praat (Boersma, 2001). For measuring the Centre of Gravity (COG) for /s/, we used the Kaldi speech recognition system for segmentation, measuring the mean of the spectral energy distribution over the segments. Each candidate for /s/ was listened to and then accepted or rejected, one by one. The COG was calculated for each of the accepted candidates.

We did not find any sign of phonetic drift over time. Interestingly, it seems that the Dutch L1 speakers had already formed different phonetic categories, since the values that they produced in the English language monologues were already clearly different, and more in line with English L1

values, than for the Dutch monologues, even for the recordings in round 1. Figure 15.3 shows this for the /s/ phoneme.

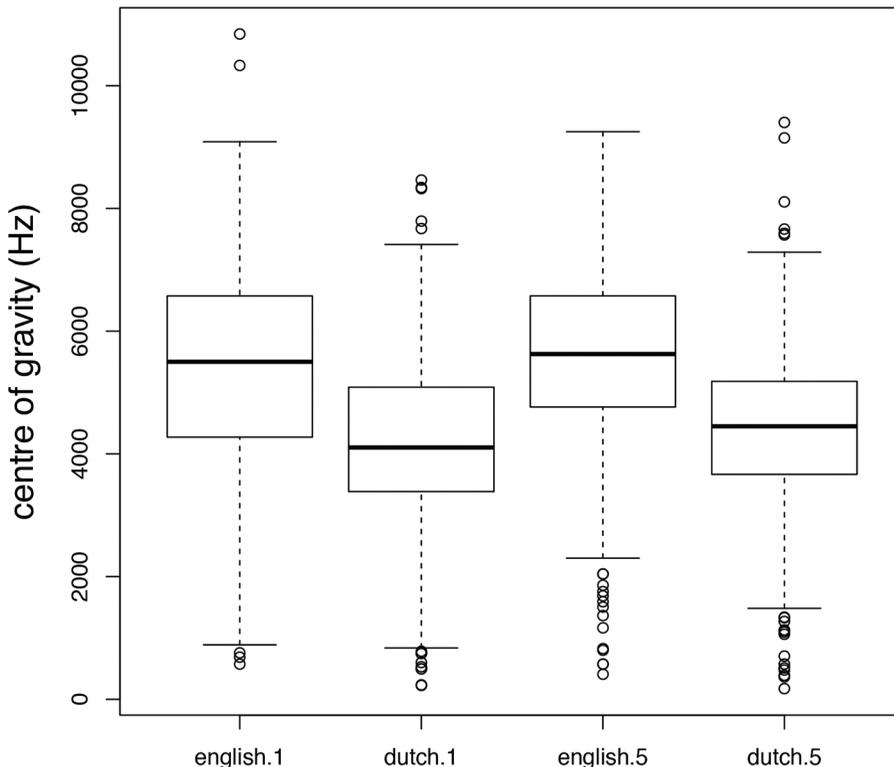
Many of the members of the L1 Dutch group had been educated in English, either at an international school in the Netherlands or abroad. It is possible that, if we isolate L1 Dutch speakers who had never been educated in English before they entered university, we may find evidence of phonetic drift. There is no clear data available on the general range of COG for /s/ in Dutch, and a comparison of this group with a similar counterpart from a regular student group in a Dutch university may provide insights into both standard values for Dutch, and a clearer view of whether this subset of our Dutch L1 group exhibit any phonetic drift for this sibilant.

For VOT values, Lisker and Abramson (1964) suggest lower values than for our group, so again, it may be worth comparing the Dutch L1 speakers from our cohort with members of the Dutch student population at large.

### 15.3.4 Further Plans for Analysis of this Corpus

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. In contrast to other models, such as the Native

## Male values for centre of gravity for /s/



**Figure 15.3:** Summary of observed Centre of Gravity (COG) values for /s/, for English and Dutch, in rounds 1 and 5 of the recordings, for Dutch L1 male speakers.

Language Magnet Model (NLM; Kuhl 1991) or the Perceptual Assimilation Model (PAM; Best 1994) that have been applied to the perception of non-native sounds, the SLM model is particularly concerned with advanced L2 learners and bilinguals, and so especially applicable to our group.

Analysis of speech from this corpus may also shed light on the nature of speaker characteristics and their dependence on which language, and with what proficiency or accent the speaker is speaking. By building acoustic speaker models using the L1 as training data for automatic speaker recognition, and testing the models on speech data from English as L2 over the three years during which a student has been recorded, we can look at whether and how the performance of the recogniser is affected as the speaker's accent changes over time towards a common variety and accent.

## References

- Best, C. T. (1994), The emergence of native-language phonological influences in infants: A perceptual assimilation model, MIT Press, chapter 6.
- Boersma, P. (2001), 'Praat, a system for doing phonetics by computer', *Glott International* 5(9/10), 341–345.
- Boersma, P. & Hamann, S. (2008), 'The evolution of auditory dispersion in bidirectional constraint grammars', *Phonology* 25(2), 217–270.
- Collins, B. & Mees, I. (2003), *The Phonetics of English and Dutch*, 5 edn, Brill Academic Publishers.
- Deterding, D. (2006), 'The north wind versus a wolf: short texts for the description and measurement of english pronunciation', *Journal of the International Phonetic Association* 36(2), 187–196.
- Evans, B. G. & Iverson, P. (2007), '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.
- Fairbanks, G. (1960), *Voice and Articulation Drillbook*, Harper and Row, chapter 5, p. 127.
- Garrett, P. (1992), '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.
- Giles, H., Coupland, N. & Coupland, J. (1991), Accommodation theory: communication, context and consequence, in J. C. H. Giles & N. Coupland, eds, 'Contexts of Accommodation: Development in Applied Sociolinguistics', Cambridge University Press, pp. 1–68.
- Goldinger, S. D. (1998), 'Signal detection comparisons of phonemic and phonetic priming: The flexible-bias problem', *Perception & Psychophysics* 60(6), 952–965.
- Jenkins, J. (2007), *English as a Lingua Franca: attitude and identity*, Oxford University Press, Oxford.
- Kuhl, P. K. (1991), 'Human adults and human infants show a perceptual magnet effect for the prototypes of speech categories, monkeys do not', *Perception & psychophysics* 50(2), 93–107.
- Lisker, L. & Abramson, A. S. (1964), 'A cross-language study of voicing in initial stops: Acoustical measurements', *Word* 20(3), 384–422.
- Orr, R., van Leeuwen, D., de Rode, J. Z. & Lohfink, G. (2015), 'L1 phonetic drift in Dutch L2 speakers of English', Paper presented at the Satellite Workshop on Phonetic Learner Corpora, Glasgow, 2015.
- Pardo, J. S. (2006), 'On phonetic convergence during conversational interaction', *Journal of the Acoustical Society of America* 119(4), 2382–2393.
- Quené, H. & Orr, R. (2014), Long-term convergence of speech rhythm in L1 and L2 English. Paper presented at the 7th Workshop on Speech Prosody, Dublin.
- Scholtmeijer, H. (1992), *Het Nederlands van de IJsselmeerpolders*, Kampen Mondiss.
- Timmis, I. (2002), 'Native-speaker norms and international English: a classroom view', *ELT Journal* 56(3), 240–249.

- Trudgill, P. (2004), *New-dialect formation: The inevitability of colonial Englishes*, Oxford University Press, USA.
- Universal Declaration of Human Rights (2013). Retrieved from <http://www.un.org/en/documents/udhr/>.
- van Wijngaarden, S. J. (1999), Speech intelligibility of native and non-native speech, in H. J. M. Steeneken, D. A. van Leeuwen & S. J. van Wijngaarden, eds, 'Multi-lingual Interoperability in Speech Technology', ISCA.
- Van Wijngaarden, S. J., Steeneken, H. J. & Houtgast, T. (2002), 'Quantifying the intelligibility of speech in noise for non-native listeners', *Journal of the Acoustical Society of America* 111(4), 1906–1916.
- Weinberger, S. (2013), 'Speech Accent Archive'. Retrieved from URL: <http://accent.gmu.edu>
- White, L. & Mattys, S. (2007), 'Calibrating rhythm: first language and second language studies', *Journal of Phonetics* 35(4), 501–522.
- Yeni-Komshian, G. H., Flege, J. E. & Liu, S. (2000), 'Pronunciation proficiency in the first and second languages of Korean–English bilinguals', *Bilingualism: Language and cognition* 3(02), 131–149.