

Phonetic similarity of /s/ in native and second language: Individual differences in learning curves

Hugo Quené

*Utrecht Institute of Linguistics OTS, Utrecht University, Trans 10, 3512 JK Utrecht,
The Netherlands
h.quene@uu.nl*

Rosemary Orr^{a)}

University College, Utrecht University, Utrecht, The Netherlands

David van Leeuwen

*Centre for Language and Speech Technology, Radboud University Nijmegen, Nijmegen,
The Netherlands
d.vanleeuwen@let.ru.nl*

Abstract: Second language learners may merge similar sounds from their native (L1) and second (L2) languages into a single phonetic category, neutralizing subphonemic differences in these similar sounds. This study investigates whether Dutch speakers produce phonetically distinct variants of /s/ in their L1 Dutch and L2 English, and whether and how this phonetic categorization develops over time. Target /s/ sounds in matching words in L1 and L2 were compared in their centre of spectral gravity. Speakers varied in their individual learning curves in the categorization of produced /s/ sounds, both in starting points and in longitudinal trajectories. After 3 years, however, all speakers had converged in producing their /s/ variants in L1 and L2 as two similar but different sounds.

© 2017 Acoustical Society of America

[RS]

Date Received: September 25, 2017 **Date Accepted:** November 8, 2017

1. Introduction

Many non-native speakers have a noticeable non-native or “foreign” accent in their second language (L2). According to the Speech Learning Model^{1–3} (SLM), speech sounds of the L2 that are phonetically similar to speakers’ native language (L1) are most difficult to pronounce in a native-like manner. Presumably, a non-native speaker tends to merge such similar L2 sounds with the corresponding L1 sounds into a single phonetic category.² In L2, the speaker will then produce a sound from this single category of L1 and L2 sounds, and will thus fail to realize the appropriate sub-phonemic differences between L1 and L2 sounds,¹ contributing to a noticeable non-native accent.

English and Dutch /s/ sounds constitute a relevant example of this distinction of two non-identical but phonetically similar sounds. Both Dutch and English /s/ are voiceless alveolar fricative consonants. But Dutch /s/ is generally produced with a more retracted tongue (towards palato-alveolar place of articulation), a flatter tongue body, and more lip rounding, as compared to English /s/.^{4,5} These articulatory differences result in a somewhat lower centre of gravity (COG) in the spectral distribution of acoustic energy during the /s/,⁶ which in turn result in the auditory impression of a less “sharp” pronunciation of /s/ in Dutch as compared to English. In articulation and in acoustics, the Dutch /s/ is somewhere between the sharper English /s/ and the duller English /ʃ/.

This paper investigates the pronunciation of Dutch (L1) and English (L2) /s/ sounds as spoken by native speakers of Dutch who are proficient in L2 English. The first question is whether these speakers produce the same or different /s/ sounds when speaking Dutch vs English. If these speakers have learned the sub-phonemic contrast described above (sounds are classified in “similar” categories per SLM), then we predict that the COG of their L1 and L2 /s/ are different. However, if the contrast is not learned (sounds are classified in a “merged” category per SLM), then we predict equal COG regardless of the language spoken.

^{a)}Deceased.

L2 speakers show considerable individual differences in their L2 proficiency. The sample of L2 speakers in the present study is relatively homogeneous (only L1 Dutch monolinguals from the upper tier of Dutch secondary education, with good proficiency in L2 English), but even these speakers may well differ due to their motivation, aptitude, and quantity and quality of exposure to English.⁷ Thus our second aim is to describe and quantify such individual differences in /s/ classification(s).

In order to investigate the possible phonetic learning of this contrast, we also tracked speakers over time (cf., e.g., Refs. 1 and 8), here in three recordings spanning a period of almost 3 years. During this time the speakers used English intensively, as a *lingua franca* in their studies and campus life, while they also used Dutch (albeit rarely while on campus). We hypothesized that speakers who had already learned the subphonemic /s/ contrast would maintain it, and that speakers who did not show the contrast in the first recording would learn it implicitly during the intervals between recordings; this hypothesis is based on speakers' general tendency to converge or accommodate towards the patterns of speech sounds of their (new) community (cf., e.g., Refs. 9–13). Thus our third question is whether and how speakers converge over time in their produced phonetic contrasts of /s/ between the two languages spoken.

2. Method

2.1 Materials

The speech materials were taken from our Longitudinal Corpus of UCU English Accents.^{14,15} This corpus contains about 1095 recording sessions of 285 students at University College Utrecht (UCU) in the Netherlands. Most of the student speakers were recorded repeatedly (maximum five “rounds” of recordings) during their 3-year stay on campus, thus allowing us to track within-speaker changes in pronunciation. Round 1 was at the onset of the first academic year, round 2 was at the end of the first year, and round 5 was at the end of the third year (recordings from intermediate rounds are not available yet). Recordings took place between August 2010 and June 2016.

Each recording session contains, among other parts, two readings of articles of the Universal Declaration of Human Rights^{16–18} (UDHR), and two spontaneous monologues of about 2 min duration on two topics of the speaker's choice, with read and spontaneous speech both in L1 Dutch and in L2 English. Speakers read the UDHR articles without problems, albeit with occasional errors in either language. In one spontaneous monologue (in each language) speakers typically described campus life, sports, or their past or upcoming travels, while in the other they talked about some recent academic work, often with the same content for the two matching monologues in the two languages. Recordings were made in a quiet furnished office, using a close-talking microphone (Sennheiser HSP 2ew), and using a FocusRite Saffire Pro 40 multichannel preamplifier and A/D converter at 44.1 kHz (exported as 16 bits linear pcm). After the first and fifth recording, speakers also filled in a questionnaire about their language background and any musical background.

Materials for the present study consisted of the /s/ segments from the UDHR readings and from the spontaneous monologues in L1 Dutch and L2 English, produced by a random sample of 25 native speakers of Dutch (19 female, 6 male; mean age 18.1 at first recording). All speakers in this sample were raised as monolinguals, had never lived in an English-speaking country, and had only started learning English at age 12 or later (mean age 13.2 at onset of learning English).

2.2 Data analysis

The /s/ segments in the speech recordings were located by the Kaldi speech recognition system¹⁹ (trained on *wsj0* and *wsj1* corpora, using *s5* recipe and *nnet2-online* configuration). Sibilant segments ($N = 8837$) varying from English /s/ to /ʃ/ were found as candidates for both Dutch and English /s/. The centre of gravity (COG) was computed for each candidate segment, using the definition of COG as used by Praat.²⁰ All candidate segments were then validated manually by the second author, yielding $N = 7384$ validated tokens of /s/; the word containing the candidate segment(s) was also transcribed. In an attempt to control for coarticulatory effects, we then selected /s/ tokens from the 24 most frequent near-homophonic and cognate words that speakers spontaneously used in *both* English and Dutch, e.g., the word *rest* (yielding $N = 538$ remaining tokens: 44 read English, 5 read Dutch, 146 spontaneous English, 343 spontaneous Dutch; with median frequency of 3 tokens per word per language). The distribution of words over speakers follows a Zipfian distribution, with a few words spoken by many speakers (*is*, *rest*, *was*) and many words spoken by only a few speakers (e.g., *festival*, *lustrum*, *west*).

The remaining /s/ tokens from these words allow for within-speaker as well as within-word comparisons between languages spoken and between rounds of recording.

The COG values of these tokens were analyzed by linear mixed-effects modeling²¹⁻²³ (LMM), with speakers as well as words as two crossed random effects. Fixed predictors were the speaker's sex (using dummy coding with females as baseline), the language spoken (L2 English vs L1 Dutch, using dummy coding with English as baseline), and the round of recording (1 or 2 or 5, using dummy coding with first recording as baseline). The interactions between language and rounds were also included. (Neither adding effects of style, read vs spontaneous, nor adding further interactions improved the model significantly, so these were excluded from the optimal model reported here.) The main effects of language and of round, and their interaction effect, were included in the random part of the model as random slopes at the speaker level,²⁴ yielding

$$\begin{aligned} \text{COG}_{i(jk)} = & \gamma_{0(00)} + \beta_{M0(00)}M + \beta_{N0(00)}N + \beta_{R20(00)}R2 + \beta_{R50(00)}R5 \\ & + \beta_{N:R20(00)}NR2 + \beta_{N:R50(00)}NR5 + u_{0(j0)} + u_{N0(j0)}N + u_{R20(j0)}R2 \\ & + u_{R50(j0)}R5 + u_{N:R20(j0)}NR2 + u_{N:R50(j0)}NR5 + v_{0(0k)} + e_{i(jk)}, \end{aligned} \quad (1)$$

where *M* denotes dummy codes for male speakers, *N* for Dutch (NL) language, *R2* for second recording, *R5* for fifth recording, *j* indicates the *j*th speaker, *k* the *k*th word, and *i* the *i*th token nested within speakers and words. This LMM acknowledges that COGs may be correlated within a speaker and within a word (due to coarticulatory effects of phonetic context), and that effects of language, round, and their interaction, may differ between speakers. Estimated fixed coefficients were evaluated by means of conservative *t* tests²⁵ taking into account the numbers of speakers, words, and fixed predictors. Fixed and random estimates were also evaluated by means of bootstrapped confidence intervals over 500 iterations.

3. Results

The fixed part of the LMM shows first, and not suprisingly, that male speakers produce significantly lower COG's than female speakers [$\beta_M = -739$, $t(17) = -2.239$, $p = 0.019$; 95% C.I. (-1443, -64)]. Second, speakers' COG is significantly lower while speaking Dutch than while speaking English, in the same phonetic contexts [$\beta_N = -1373$, $t(17) = -4.429$, $p < 0.001$; 95% C.I. (-1958, -733)]. Figure 1 shows that speakers' COGs tend to fall below the (dotted) $y = x$ diagonal, i.e., COGs tend to be lower in Dutch than in English; this tendency is captured by the dashed LMM regression line.

Relative to the 1st recording, average COGs were slightly higher in the second recording, but not in the fifth recording, yielding a non-significant effect of rounds [$F(2, 17) = 2.99$, $p = 0.077$]. None of the interaction effects were significant. Thus the overall contrast in COG of English /s/ (estimated mean 5778 Hz) and Dutch /s/

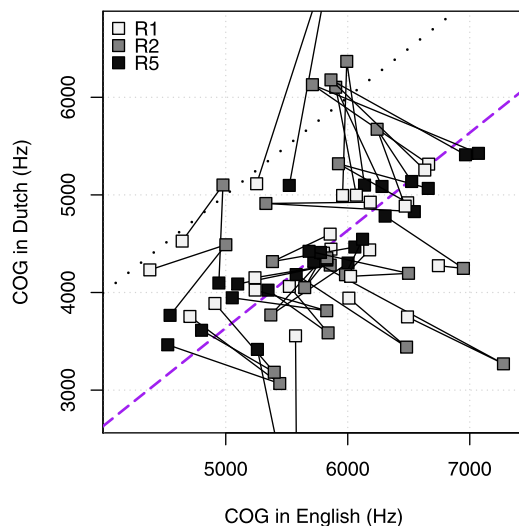


Fig. 1. (Color online) Estimated spectral centre of gravity (COG) in Hz, in English and in Dutch, for each speaker and recording separately. The black dotted diagonal indicates equal COG in both languages ($y = x$), the colored dashed line indicates the estimated relation between English and Dutch COG values across recordings. For clarity, two outlier points at coordinates (5584, 1459) and (5880, 7542) are not plotted.

(estimated mean 4405 Hz) does not change significantly over time across all speakers collectively (but see below).

Figure 1 illustrates that in the 1st recording (lightest symbols), most speakers produce a lower COG in L1 Dutch than in L2 English. Apparently, these speakers have already acquired the phonetic difference between English and Dutch /s/ before their arrival on campus. Three speakers, however, still produce approximately equal COG's in both languages (near the dotted diagonal). In the second recording (grey symbols), average COG's have drifted for most speakers. Some speakers produce very high COG in their English (overshoot, rightward shift in Fig. 1), and others produce /s/ with a typically "English" COG also when speaking Dutch (upward shift). In the fifth recording (darkest symbols), all speakers have converged towards the dashed regression line which indicates the overall phonetic contrast in COG between L1 and L2 variants of /s/. Almost all speakers consistently use the appropriate variant of /s/ in both languages.

These patterns are corroborated by the random part of the LMM, which contains random intercepts for speakers [using English tokens of first interview as baseline, $\sigma(u_{0(j0)}) = 871$ Hz, bootstrapped 95% C.I. (440, 1436)], as well as for words [$\sigma(v_{0(0k)}) = 485$ Hz, bootstrapped 95% C.I. (0, 893)]. Moreover, the between-speaker standard deviation remains approximately equal in the second interview [English, $\sigma(u_{R2}) = 889$, (272, 1985)], whereas it is somewhat smaller in the fifth interview, but not significantly so [English, $\sigma(u_{R5}) = 468$, (236, 1406)]. Figure 1 confirms that the between-speaker variability in L2 English (along horizontal axis) is about the same for all three recordings (shades).

In the first recording, the between-speaker standard deviation in L1 Dutch tokens is about equal [Dutch, $\sigma(u_N) = 889$, (324, 1542)] to between-speaker variability in English. In the second recording, however, variation between speakers has increased considerably [Dutch, $\sigma(u_{N:R2}) = 1311$, (585, 2677)], whereafter it decreases again in the fifth recording [$\sigma(u_{N:R5}) = 782$, (279, 1834)], as illustrated in Fig. 1.

These random coefficients confirm that between-speaker variability in the COG of /s/ remains about the same across the L2 English parts of the interviews. This suggests that the speakers' English accent is relatively stable, as far as the /s/ pronunciation is concerned, and that it does not change during the years of intensive usage of English during which the interviews were collected. Speakers' accents in their L1 Dutch, by contrast, are less stable. Some speakers enhance the contrast in their L1 Dutch (by moving "downward" below the dashed regression line in Fig. 1, "overshoot"), or they use the sharper L2 English /s/ in their L1 Dutch too (moving "upward"), or they use the duller Dutch /s/ in their L2 English too (moving "leftwards"). At the end of their first year of study, speakers' L1 Dutch is clearly affected, by speakers' intensive and near-exclusive use of English as the lingua franca on campus, so that speakers vary more in their L1 accents of /s/ than in their L2 accents. However, by the time of the fifth recording, after three years of study, all speakers have learned to produce the same contrast (near lower diagonal) appropriate for this peer group. They have all learned to produce the appropriate /s/ for this group of speakers, while they have followed different developmental routes towards this equilibrium.

4. Discussion

The results of this within-speaker, longitudinal corpus study corroborate the acoustic difference between Dutch and English /s/.^{4,5} The native Dutch speakers in the current study were sampled from a highly selective undergraduate college in the Netherlands. In order to be admitted to this particular college, prospective students from a Dutch background need to have received at least 6 years of secondary education in English, and they need to show very good grades for English in their secondary education (at least 8/10 points). With speakers thus sampled from the top end of their peers' proficiency in L2 English, it may not come as a surprise that most of these proficient L2 English speakers had already learned the sub-phonemic contrast in L1 and L2 realization of /s/ before their first recording, even though it probably was never taught explicitly. This confirms similar reports of relatively large L1–L2 contrasts in young, proficient speakers: early L1 Spanish learners of L2 English showed a larger contrast between their L1 Spanish and L2 English stop consonants than late learners,²⁶ and young Korean–Mandarin bilinguals showed a larger contrast in COG between their Korean and Mandarin sibilants than older L1 Korean L2 Mandarin speakers.²⁷

The higher COG for /s/ in L2 English than in L1 Dutch corresponds with a more forward tongue position in producing English /s/ than in Dutch /s/. Similar results have been reported for vowels too, with higher formants for English vowels

than for Dutch vowels, again corresponding to a more forward tongue position in producing English vowels as compared to Dutch vowels.²⁸ Together these findings suggest that speakers may have noticeably different articulatory-phonetic settings for Dutch and English, with a more forward tongue position in English than in Dutch, which affects both vowels and consonants in a similar way. More generally, these findings confirm that articulatory-phonetic settings may be established from acoustic features, and that these settings are relevant not only to compare native speakers across languages, but also to compare across languages spoken by the same speaker.²⁹

Speakers in the current study were monolingual native speakers of Dutch, educated in the same academic tier of the Dutch school system. In spite of this homogeneity in the sample of speakers, some remarkable individual differences were observed in the first recording. The expected contrast in COG of /s/ was present in the fixed (general) part of the LMM, and this contrast was indeed found for 22/25 speakers. However, the *random* part of the LMM shows that 3/25 speakers did not show this contrast in their first recording, and that only two of these have *learned* the contrast between their first and fifth recording.

Moreover, 2/22 speakers produced a very *large* contrast in their first recording, but then later *reduced* the contrast. Neither the latter fraction of two speakers, nor the former fraction of three speakers, could be uniquely associated with any of the speakers' answers to questions in the entry or exit questionnaires. Thus, the individual differences in speakers' learning curves might be associated with latent traits not uncovered by our questionnaires, or these differences might be entirely random. Taking these individual differences into account in the LMM, in the form of random slopes, improved the model fit considerably. This suggests that in studying speech production and speech perception, findings should be analysed and reported not only in terms of overall patterns across participants, but also in terms of individual differences among participants.^{7,30}

The results of the 1st recording corroborate previous reports about different articulations of similar /s/ sounds in English and in Dutch. Before commencing their undergraduate studies, most speakers in this study have already learned to produce the appropriate variant of the "similar" sound in each language, at least in the 24 selected cognate words. Thus, speakers seem to employ two distinct articulatory settings for English and for Dutch, at the time of their first recording.

Moreover, results of the second and fifth recording indicate that speakers' accents in L1 and L2 remain plastic well into adolescence, as predicted by the SLM. After about 9 months of intensive usage of English as a lingua franca (in the second recording), many speakers' accents had diverged into multiple directions. Remarkably, some speakers had even "unlearned" the articulatory difference in /s/ between the two languages. Speakers' accents remained plastic in subsequent years too: in the last recording, accents had drifted again, so that all speakers had converged to two similar but phonetically different variants of /s/ in their Dutch and English speech.

Acknowledgments

Data for this study were collected by the first two authors, and processed and analyzed by all three authors, before the death of the second author in November 2016. Preliminary findings were presented at a workshop on Phonetic Learner Corpora (Glasgow, 12 August 2015) and in Ref. 14. We thank Martin Everaert, Hans Van de Velde, Rias van den Doel, Jocelyn Ballantyne, Frank Wijnen, Jürgen Trouvain, and Willemijn Heeren for support and discussions, and we thank all UCU student speakers and facilitators for their help in collecting data.

References and links

- ¹J. E. Flege, "The production of 'new' and 'similar' phones in a foreign language: Evidence for the effect of equivalence classification," *J. Phon.* **15**, 47–65 (1987).
- ²J. E. Flege, "Second-language speech learning: Theory, findings, and problems," in *Speech Perception and Linguistic Experience: Issues in Cross-language Research*, edited by W. Strange (York Press, Timonium, MD, 1995), pp. 229–273.
- ³J. E. Flege, "Language contact in bilingualism: Phonetic system interactions," in *Laboratory Phonology 9*, edited by J. Cole and J. I. Hualde (Mouton de Gruyter, Berlin, 2007), pp. 353–381.
- ⁴B. Collins and I. M. Mees, *The Phonetics of English and Dutch*, 5th ed. (Brill, Leiden, 2003).
- ⁵M. Wieling, P. Veenstra, P. Adank, A. Weber, and M. Tiede, "Comparing L1 and L2 speakers using articulatory data," in *Proceedings of the International Congress of Phonetic Sciences*, Glasgow (2015).
- ⁶K. N. Stevens, *Acoustic Phonetics* (MIT Press, Cambridge, MA, 1998).
- ⁷S. M. Gass and A. Mackey, eds., *The Routledge Handbook of Second Language Acquisition* (Routledge, London, 2012).

- ⁸M. Milenova, “The acquisition of new and similar sounds: Evidence from Bulgarian learners of modern Greek,” in *11th International Conference on Greek Linguistics* (2015).
- ⁹H. Giles, N. Coupland, and J. Coupland, “Accommodation theory: Communication, context, and consequence,” in *Contexts of Accommodation: Developments in Applied Sociolinguistics*, edited by H. Giles, J. Coupland, and N. Coupland (Cambridge University Press, Cambridge, 1991), pp. 1–68.
- ¹⁰H. Scholtmeijer, *Het Nederlands van de IJsselmeerpolders (The Dutch Language of the IJsselmeerpolders)* (Mondiss, Kampen, 1992).
- ¹¹J. S. Pardo, “On phonetic convergence during conversational interaction,” *J. Acoust. Soc. Am.* **119**, 2382–2393 (2006).
- ¹²B. G. Evans and P. Iverson, “Plasticity in vowel perception and production: A study of accent change in young adults” *J. Acoust. Soc. Am.* **121**, 3814–3826 (2007).
- ¹³A. Nardy, J.-P. Chevrot, and S. Barbu, “Sociolinguistic convergence and social interactions within a group of preschoolers: A longitudinal study,” *Lang. Var. Change* **26**, 273–301 (2014).
- ¹⁴R. Orr and H. Quené, “D-LUCEA: Curation of the UCU Accent Project data,” in *CLARIN in the Low Countries*, edited by J. Odijk and A. van Hessen (Ubiquity, London, in press), pp. 177–190.
- ¹⁵“LUCEA: Longitudinal corpus of University College Utrecht English accents,” <http://lucea.wp.hum.uu.nl> (Last viewed 24 September 2017).
- ¹⁶“The Universal Declaration of Human Rights” (1948/2010), <http://www.ohchr.org/EN/UDHR/Pages/UDHRIndex.aspx> (Last viewed 24 September 2017).
- ¹⁷A. R. Bradlow, L. Ackerman, L. A. Burchfield, L. Hesterberg, J. Luque, and K. Mok, “Language- and talker-dependent variation in global features of native and non-native speech,” in *Proceedings of the XVII International Congress of Phonetic Sciences*, Hong Kong (17–21 August 2011), pp. 356–359.
- ¹⁸“ALLSTAR: Archive of L1 and L2 scripted and spontaneous transcripts and recordings,” http://groups.linguistics.northwestern.edu/speech_comm_group/allstar (Last viewed 24 September 2017).
- ¹⁹D. Povey, A. Ghoshal, G. Boulianne, L. Burget, O. Glembek, N. Goel, M. Hannemann, P. Motlicek, Y. Qian, P. Schwarz, J. Silovsky, G. Stemmer, and K. Vesely, “The Kaldi speech recognition toolkit,” in *IEEE 2011 Workshop on Automatic Speech Recognition and Understanding* (2011), Vol. CFP11SRW-USB.
- ²⁰P. Boersma and D. Weenink, “Praat: Doing phonetics by computer,” <http://www.praat.org> (Last viewed 12 July 2016).
- ²¹H. Quené and H. Van den Bergh, “Examples of mixed-effects modeling with crossed random effects and with binomial data,” *J. Mem. Lang.* **59**, 413–425 (2008).
- ²²D. Bates, M. Mächler, B. Bolker, and S. Walker, “Fitting linear mixed-effects models using lme4,” *J. Stat. Softw.* **67**, 1–48 (2015).
- ²³R Core Team, “R: A language and environment for statistical computing,” R Foundation for Statistical Computing, Vienna, Austria, <https://www.R-project.org> (Last viewed 9 June 2017).
- ²⁴T. Snijders and R. Bosker, *Multilevel Analysis: An Introduction to Basic and Advanced Multilevel Modeling* (Sage, London, 1999).
- ²⁵J. J. Hox, *Multilevel Analysis: Techniques and Applications*, 2nd ed. (Lawrence Erlbaum, Mahwah, NJ, 2010).
- ²⁶J. E. Flege and W. Eefting, “Production and perception of English stops by native Spanish speakers,” *J. Phon.* **15**, 67–83 (1987).
- ²⁷J. Schertz, Y. Kang, and S. Han, “Cross-language correspondences in the face of change: Phonetic independence versus convergence in two Korean-Mandarin bilingual communities,” *Int. J. Bilingual.*, in press (2017).
- ²⁸S. Bultena, “Are you in English gear? A study on the acoustic correlates of articulatory settings in English and Dutch,” *Toegepaste Taalwetenschap Artikelen* **79**, 9–20 (2008).
- ²⁹I. Mennen, J. M. Scobbie, E. De Leeuw, S. Schaeffler, and F. Schaeffler, “Measuring language-specific phonetic settings,” *Sec. Lang. Res.* **26**, 13–41 (2010).
- ³⁰H. Quené, “Multilevel modeling of between-speaker and within-speaker variation in spontaneous speech tempo,” *J. Acoust. Soc. Am.* **123**, 1104–1113 (2008).