

Una Cunningham

Högskolan Dalarna, Sweden

ACOUSTIC VARIABILITY IN THE PRODUCTION OF ENGLISH VOWELS BY NATIVE AND NON- NATIVE SPEAKERS

1 What is the difference between native and non-native speech?

1.1 Introduction

The features of non-native speech which distinguish it from native speech are often difficult to pin down. It is possible to be a native speaker of any of a vast number of varieties of English. These varieties each have their phonetic characteristics which allow them to be identified by speakers of the varieties in question and by others. The phonetic differences between the accents represented by these varieties are very great. It is impossible to indicate any particular configuration of vowels in the acoustic vowel space or set of consonant articulations which all native-speaker varieties of English have in common and which non-native speakers do not share. The differences within the group of native speakers are often as large or larger than the differences between native and non-native speakers. And yet many studies have been carried out demonstrating that it is not very difficult to distinguish between native and non-native speech, e.g. the classic work of Brennan and Brennan (1981). Native speakers are generally quite good at “spotting the non-native speaker”, although they can sometimes be misled by particular phonetic features which strongly suggest nativeness (glottaling and Estuaryesque vowel fronting come to mind as candidates here).

1.2 Vowel quality and quantity

Differences in vowel quality are a major part of what distinguishes one variety of English from another. Systematic, phonological differences are crucial for the difference between RP and GenAm or Irish English, and more subtle phonetic differences in vowel quality explain how we can tell GenAm from Canadian, Birmingham from Liverpool, or Ulster Scots from Glasgow or maybe even General Australian from New Zealand English or RP. Wells' (1982) using his set of key words (KIT, etc.) offers a neat summary of vowel differences between many varieties. His system allows the potential contrasts offered by the English language to be easily described and discussed, although not with much phonetic detail.

Vowel quantity is less obviously useful as a way for listeners to distinguish varieties in real time, although there certainly are well-known systematic differences in the temporal relationships found in different varieties. It may well be the case that some accents make extensive linguistic use of vowel length distinctions for phoneme identification (e.g. in *beat-bit*). Others may rely less on temporal relationships and more on vowel quality in such situations. This variation is not well documented as far as I am aware. Certainly this is a potential area where native and non-native pronunciation may differ from each other.

One of the better-documented features of native English pronunciation is the use of vowel length differences as the most salient (to native listeners) cue to post vocalic consonant voicing (e.g. Gimson 1974). The enhancement of fortis clipping for linguistic purposes is an excellent candidate for being overlooked or otherwise misinterpreted by non-native speakers of English. This is a feature of some kinds of non-native accents which gives rise to significant comprehensibility issues, and may well be a useful cue for native listeners' identification of a speaker as non-native.

1.3 Consonant articulation and quantity

The exact articulation of consonants is obviously a potentially important factor for the identification of non-native accents. Influences from the L1, on the one hand, and the typological rareness of certain consonant articulations (e.g. [D, T]), on the other, may leave the non-native speaker of English with a particular set of consonants which only partly resembles that of any native speaker, regardless of variety. But of course the matter of learner models cannot be disregarded in this respect. Even if the non-

native speaker does not aspire to the target of a standard (or indeed non-standard) native variety, they have presumably had a native variety (generally GA or RP) as a model for their production throughout their formal language education. These accents of English have largely the same consonant articulations, phonotactic constraints and consonant quantity conditions. While consonant articulations may seem less salient as identifiers of non-native pronunciation than vowel quality, it has been shown that they can be give-aways. Flege and Hammond (1982), Flege (1984), showed that 30ms of speech was enough for American English-speaking listeners to distinguish native English from French-accented speech. He found that listeners were particularly sensitive to the noise burst (and presumable formant transitions and presence/absence of aspiration) around the release of initial /t/.

The exact place of consonant articulation is another variable that is not constant for all native speakers, but the standard models do have e.g. alveolar stops where other languages have dental stops, and a lot of freedom for coarticulatory assimilation to take place in the relatively empty velar area, which is certainly not the case for all languages. Similarly, manner of articulation can cause difficulty for some non-native groups, especially in the case of the affricates of English. The exact degree of tongue grooving and place of maximum constriction for /s/, the lip-rounding and colour of the frication of /ʃ/ or the articulation of the notoriously difficult-to-grasp British English /t/ can quite literally be shibboleths.

Phonotactics are one of the most problematic areas for non-native speakers. English is an unusually permissive language in this respect (although Polish does outclass English here, c.f. Jassem 2003). The clusters allowed in citation style speech are extensive, and in casual spontaneous speech there are even more possibilities. Speakers of other languages do not easily free themselves of the phonetic constraints of their L1. An additional effect here is in L1 phonological rules which are often difficult for speakers to consciously access and still less to suppress when speaking English. The phonotactic difficulties of non-native speakers are in many cases a very prominent feature of an accent.

1.4 Timing relationships

The relationship between the durational conditions of certain phonetic elements is subject to systematic manipulation. In English, for example, the durations of vowels and of the stops which follow them have an

inverse durational relationship. In Swedish this relationship is enhanced and is used linguistically as a salient cue to phoneme identification. Sometimes VC timing relationships are linguistically significant, at other times the result of the application of a phonological rule. In any case, this is an area where different L1s will work differently. Some languages, e.g. Swedish, Italian or Estonian, make heavier use of timing relationships than other languages, such as e.g. Spanish (c.f. McAllister et al. 2002) and it is reasonable to expect that this will be reflected in the English of speakers of these languages.

In a Swedish study of an impersonator's success in matching the phonetic patterns of the person being imitated it was found that it was a lot easier to accurately trace the formant frequencies and frequency means (Eriksson and Wretling 1997) than the timing pattern (Wretling and Eriksson 1998). They suggest in their 1998 paper that since speech is an "automated motor activity" it is likely that timing patterns in speech are "fairly stable within a speaker". This could suggest that this will be a particularly difficult area for second language learners, which has indeed been found in many studies (for example, Flege 1984, Cunningham 1986, Cunningham-Andersson 1987, 2003). But the non-observance of the timing relationships prescribed in the standard pronunciation models may not be a very salient cue to non-nativeness in the case of English.

1.5 Variability

It is sometimes taken for granted that non-native speakers are more variable in their production than native speakers (e.g. Oh et al. 2007, Jongman and Wade 2007). It seems clear that this is the case, for example, in the observance of grammatical gender for English L1 Swedish. Also in, for example, the observation of subject-verb concord in the English of Swedish L1. Sometimes a non-native speaker/writer may get it right, sometimes they may not. These errors are not likely to show up in native speakers' production (at least if they speak and write a standard variety of English) as slips. It is less clear that this will be the case for pronunciation targets. Native speakers also have variable production, particularly if they vacillate between standard and non-standard pronunciations, such as in Trudgill's classic Norwich studies (Trudgill 1974) where he found that almost all his informants had some tokens of [ɪn] and some of [ɪn̩] for the *-ing* suffix. Consider the production of /t, d/ as (inter)dental fricatives. Native speakers of standard varieties may almost always have a fricative articulation of these, but in some contexts, due to casual speech

phenomena the articulation can be changed. Non-native speakers, on the other hand, may be able to produce these sounds as (inter)dental fricatives when they are able to pay attention to the pronunciation of the speech they are producing, such as in citation form speech with wordlists, but when they are speaking spontaneously there might well be alternative articulations, such as [s, z] or [t, d]. In speech recognition work there appears to be an assumption that non-native speakers are more variable in their articulations than native speakers (e.g. Oh et al. 2007). This variability, which is presumably an expression of uncertainty and lack of familiarity with target language articulations, may well be a factor which can be used by native listeners to identify non-native speakers, although the occasional non-native pronunciations are probably more salient than the occasional native-like pronunciations.

1.6 Accentedness

Non-native speech can be more or less accented. The concept of accentedness can be difficult to pin down. Cunningham-Andersson and Engstrand (1989) found that phonetically naïve listeners were able to estimate global accentedness on a scale in a way that correlates strongly with “expert” phonetic judgement on a variety of parameters. Strong non-native accent will generally interfere with intelligibility. An accent can be perceived as stronger if it has more non-native features, or if the non-native features it has are more pronounced.

2 Method

2.1 Informants

There are four main groups of informants in this study:

- Group one is selected from a longitudinal corpus of ten Swedish students attending an English-medium upper-secondary programme at a Swedish school. The students were sampled in their first and final (sixth) semester. This study uses part of the data from the English material.
- Group two is an embryonic collection of Hiberno-English voices – the beginnings of a corpus of women in South Tyrone. The women whose data are presented here are 17, 70 and 80 years old.

- Group three is a sample of convenience, composed of assorted native speakers (RP, London, GA, California, North and South Hiberno-English) I found in my immediate vicinity at the English Department of Högskolan Dalarna.
- Group four is made up of Vietnamese women (about 20-40 years old), some of whom are university teachers of English and speak English with moderate Vietnamese accents, others who are teachers of other subjects or administrators, and speak English with stronger Vietnamese accents.

2.2 Material

The material used in this study was the same for all informants – a text and a word list (see Appendix) containing a selection of the words from the text, chosen with the intention of eliciting tokens of all vowel phonemes in several contexts. The text was read first, and then the word list was read twice by each speaker. During other analyses of parts of the material from Group one (Cunningham 2004) it became clear that the English vowel /u:/ was subject to some variation. In particular a few of the students tended to pronounce this vowel with a raised F_2 (a phenomenon known as fronting) in the context of the word *choose*. Now the reasons for this can be many – the fronting could be conditioned by the context, in particular the palatoalveolar initial consonant. Also this kind of fronting is increasingly heard in young British speakers (noted even by Gimson 1974:120 as “considerable centralization”) and is also a variable with sociolectal significance in a number of American studies discussed by Cheshire (2002). Perhaps these young speakers, who have extremely high integrative motivation, have latched onto this nativeness marker.

3 Results and discussion

3.1 Variability in non-native speech

The formant frequencies of the high vowels in the wordlist were the subject of this study. The words shown in Table 1 were analysed as they occurred in the wordlist.

Table 1 stimulus words with high vowels.

<i>/i:/</i>	<i>/I/</i>	<i>/u:/</i>	<i>/U/</i>
<i>sheep</i>	<i>grin</i>	<i>pool</i>	<i>could</i>
<i>believe</i>	<i>still</i>	<i>school</i>	<i>room</i>
<i>trees</i>	<i>think</i>	<i>through</i>	<i>pull</i>
<i>green</i>	<i>this</i>	<i>choose</i>	<i>would</i>
<i>see</i>	<i>quickly</i>		
<i>feel</i>	<i>ship</i>		
<i>leaves</i>	<i>window</i>		

The formant measurements were made using WaveSurfer at a point of the vowel where formant frequencies were judged to be relatively steady and representative.

Figure 1 shows the high vowels of the speaker we call Sara, one of the Swedish speakers of English in group one at the first recording (when they were 16 years old and at the beginning of their 3-year English-medium programme) as a formant plot with F_1 plotted against F_2 arranged in such a way as to resemble the articulatory vowel quadrilateral. Sara can be seen to have discrete categories for her */i:/*, */I/* and */U/* except that her */u:/* spreads out over the part of the vowel space occupied by her */I/* and */U/*. Now vowel quality is not the only way open for speakers to distinguish vowels, and it may well be that Sara maintains a clear quantity distinction between */u:/* and */I/* and */U/*. Notice the two tokens of */u:/* that are farthest to the front, with a F_2 value above 2000Hz. These are the two tokens of *choose*.

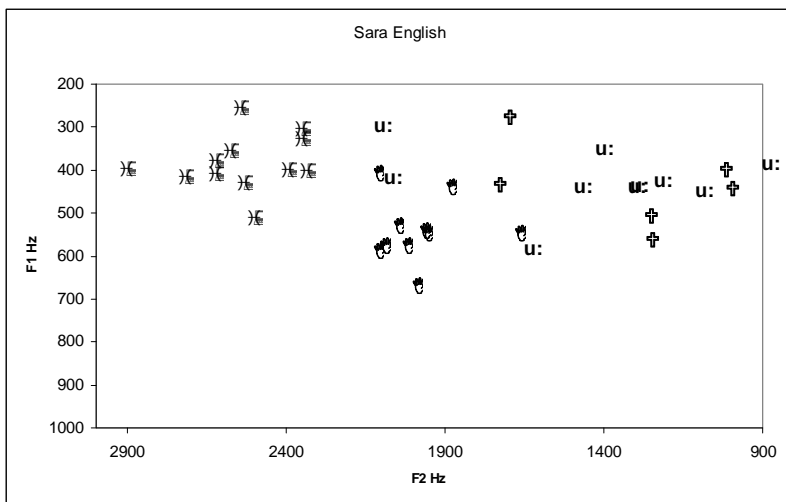


Fig. 1 Sara's high vowels in English

3.2 Variability in native speech

Let us examine our notion that non-native speakers are more variable in their pronunciation than native speakers, by comparing the native Swedish speech and the non-native English speech of the same speakers. Jongman and Wade (2007) looked at acoustic variability in vowel quality for native speakers of English and Spanish-L1 speakers uttering the same eight English vowels, and found considerable variability and overlap in the vowel qualities of the non-native speakers. Figure 2 compares the same speaker, Sara, speaking Swedish.

As can be seen here, Sara has a lot of variability in her native language Swedish in the case of the /ɪ/ vowel in the Swedish words *skinn*, *flicka*, *vitt* and the /ʊ/ in *gubbe*, *full*, *skulle* (each word uttered twice in citation form). Presumably this is caused by coarticulatory effects, e.g. the /k/ following the vowel in *flicka* causing the [ɪ] to have a lower F₂. Notice that her /i:/ is nicely gathered, but with a lowered F₂. This centralisation or backing is a socioeconomically significant feature of the pronunciation of the middle class in the region round Stockholm (known in Swedish as *Viby-i* or *Lidingö-i* after places where this sound is said to be particularly frequent) which is certainly not reflected in Sara's English.

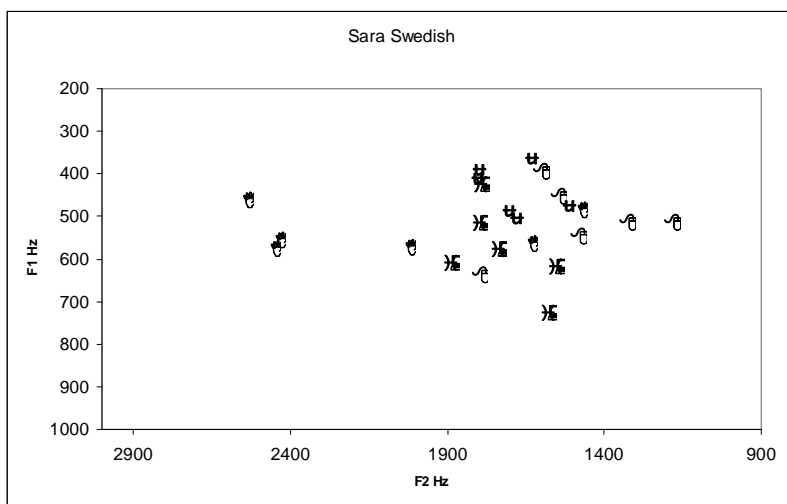


Fig. 2 Sara's high vowels in Swedish

3.3 The case of school

To make the picture clearer, and to eliminate the effects of phonetic context on vowel quality, the /u:/ in *school* was studied in detail. The pronunciation of this vowel will be affected by the context, perhaps particularly by the /l/ following the vowel which will, in many native speakers' speech, be velarised or even vocalized as [U]-like. Figure 3 shows the vowel quality of five tokens of the vowel in *school* as uttered in twice in citation form and in three times a read text by a selection of native speakers (group three above).

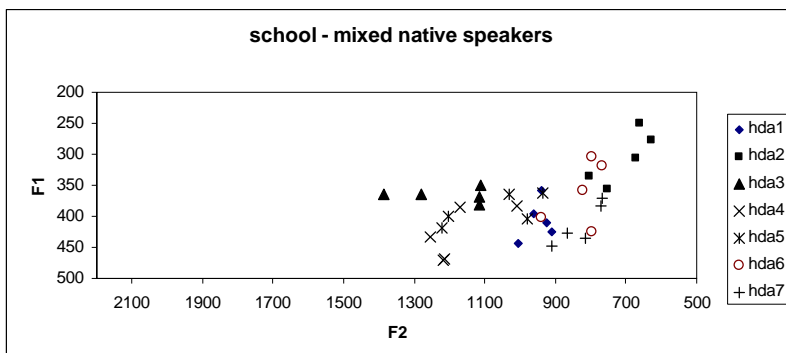


Fig. 3 Native speaker pronunciations of /u:/ in school

The speakers here are different from each other in the acoustic quality of their vowels, but there is not a great deal of within-speaker variation. Interestingly, the geographical varieties represented are quite distinct. HDA1 and HDA6 are female speakers of GenAm, and they have the lowest F₂ after HDA2 who is a male RP speaker. HDA7 (female) from London is also represented. The three speakers with the highest F₂ are HDA5 from Australia, HDA4 from Northern Ireland and HDA3 from the Republic of Ireland (all female).

Compare these native speakers with three speakers from a single accent of English, that of South Tyrone in Northern Ireland, as shown in figure 4 below.

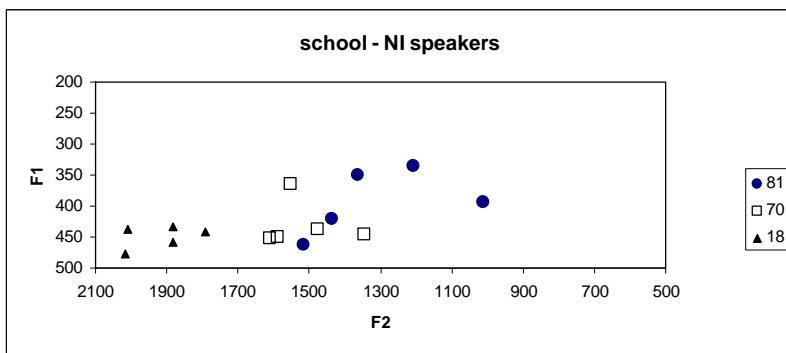


Fig. 4 Northern Hiberno-English women

Notice that the 81-year old speaker shows more variation in the quality of her vowel than the other two speakers. Notice also that the youngest woman, just 18-years old, seems to be most “extreme” in her pronunciation, i.e. her F2 is much higher than any other speaker, perhaps because she has never lived outside the area of rural South Tyrone.

3.4 Learner varieties

So, we have perhaps established that some native speakers have quite a lot of within-speaker variability, and the between-speaker variability is considerable. But let us then consider non-native speakers, again in the case of the word *school*, as it occurs three times in the text and twice in citation form in the wordlist.

Figure 5 shows relatively proficient speakers of Vietnamese-accented English (university teachers of English) and Figure 6 shows less proficient speakers (university administrative staff).

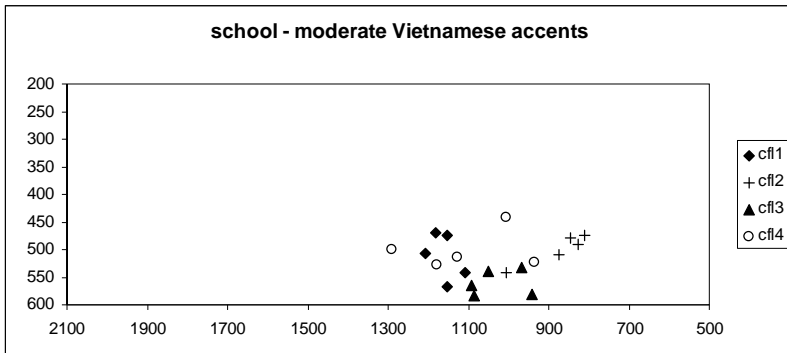


Fig. 5 Moderately Vietnamese-accented English

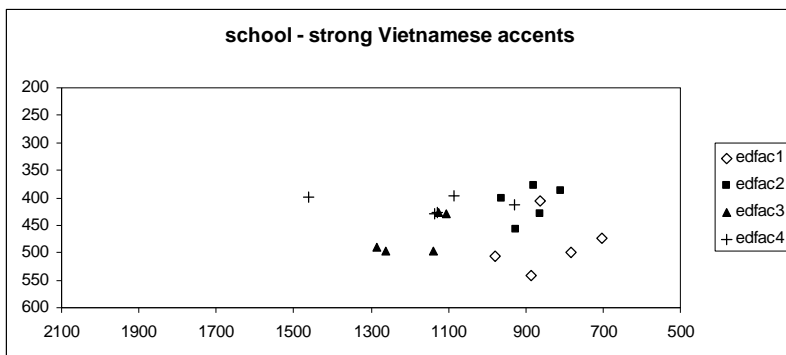


Fig. 6 Strongly Vietnamese-accented English

Now here there does appear to be an effect. The more proficient learners have less within-speaker and less between-speaker variation than the less proficient speakers over the five tokens of the word *school*. But of course there may be other explanations for this kind of difference. We cannot be sure that individual variation is not behind this apparent difference. So let us then see what happens in a single speaker over time. Figure 7 shows the pronunciation of the vowel in *school* was compared in a single speaker from the longitudinal study of the Swedish girls in group one in the first and last of the six-semester (three-year) course.

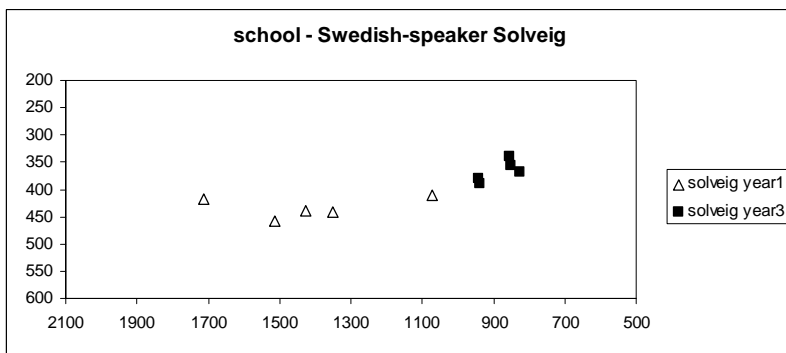


Fig. 7 Solveig: /u:/ in school at the beginning and end of a 3-year English-medium programme

Notice that the speaker we call Solveig has considerably more variable pronunciation at the beginning of her three years of study than at the end. Her /u:/ is very concentrated by year 3 in a area of the acoustic vowel

space towards that occupied by JWH, our RP speaker (F_1 between 250 and 350 Hz and F_2 between 600 and 800 Hz, marked by a circle in Figure 7).

Figure 8 shows the equivalent measurements for Susanna, another speaker in group one.

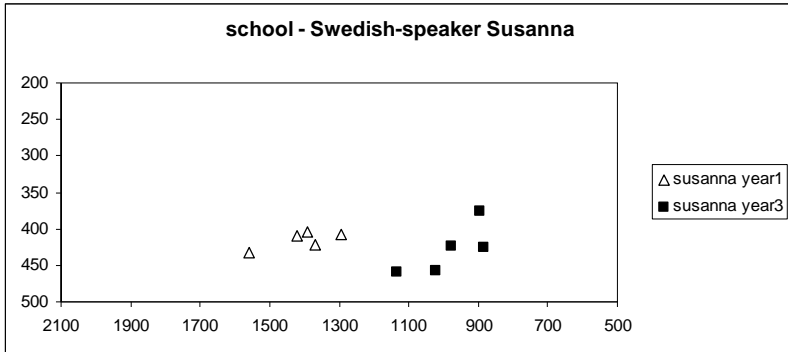


Fig. 8 Susanna: /u:/in school at the beginning and end of a 3-year English-medium programme

Here again we can see that the pronunciation has moved in the direction of RP /u:/, though the year three pronunciation cannot be said to demonstrate less within-speaker variation than the year one pronunciation. As a point of comparison, let us consider if a parallel development is happening in Swedish high back vowels of these speakers. Figure 9 shows Solveig's Swedish pronunciation of the /u:/ in the words *bot* and *bod*, each occurring once in the wordlist and once in the text.

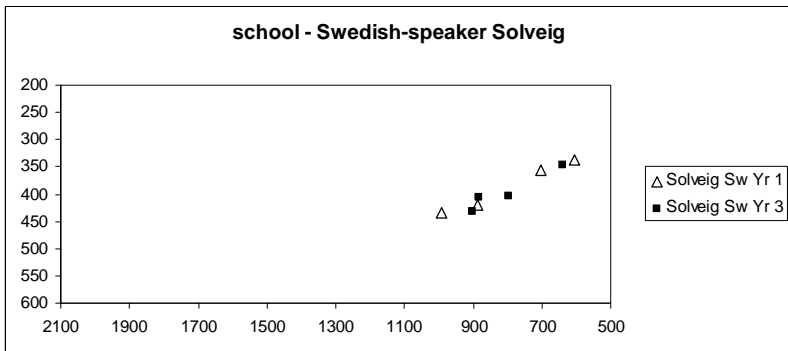


Fig. 9 Solveig: /u:/ in *bot* and *bod* at the beginning and end of a 3-year English-medium programme

Clearly, very little has changed in Solveig's Swedish. She still has some variability in her Swedish /u:/, but the acoustic quality of the vowels is still in the same area of the vowel space in year three as it was in year one.

4 Conclusion

The main conclusion that can be drawn from the results above is that the amount of variability shown by a speaker is very individual. Some native speakers seem to vary a lot in their vowel quality. Perhaps this is due to their being drawn by more than one regional or sociolectal variety. Non-native speakers also seem to be more or less variable, but here we do seem to see an effect of increasing proficiency, in that the more proficient (or perhaps more schooled) Vietnamese and Swedish speakers of English appear to be less variable than the less proficient or less schooled speakers. This is a thread that merits further attention, and obviously a larger set of words.

The status of variability as a potential cue to nativeness or even to non-nativeness is far from clear. The variety of vowel quality demonstrated by those who call themselves native speakers of English is enormous. It is not at this stage possible to draw conclusions on variability, but this is certainly an area of second language speech that will yield interesting results in further work.

Appendix - Stimulus material

Wordlist

sheep because this believe boy choose comfort could ship day pull shut
small still think become grin through trees see very longer green room feel
country would places window school like leaves unhappy great thought
quickly adult house friends man pool run govern high

Text

A small boy lives in this house. There are fields with sheep all round the house. His room is at the back and he can see his school from the window through the green leaves of the trees if he wants to pull them to one side.

He feels very unhappy because he has no friends and he believes that if he could become adult quickly he wouldn't have to go to school. If he could choose, he would like to govern the country and think great thoughts about the world and have friends in high places. But he is not yet a man and he must still shut up and do what he is told.

One day he might run away from school and make his way to another country in a ship. But really, it is not long until he will no longer be a boy. He can comfort himself with that thought. He starts to grin and goes down to the pool for a swim.

References

- Brennan, E. and Brennan, J. (1981). Measurement of accent and attitude toward Mexican American speech. *Journal of Psycholinguistic Research* **10**, 487-501.
- Cheshire, J. (2002). Sex and gender in variationist research. In J. K. Chambers, Peter Trudgill, Natalie Schilling-Estes (eds.) *The Handbook of Language Variation and Change*. Malden, MA : Blackwell Publishers, 423-443.
- Cunningham, U. (1986). *A Linguistic Theory of Timing*. PhD. Dissertation. Dept. of Linguistics, University of Nottingham.
- Cunningham-Andersson, U. (1987). Durational correlates of post-vocalic voicing in English spoken by English and Spanish speakers. Papers from the Swedish Phonetics Conference in Uppsala, 17-18 October 1986. RUUL 17, 87-92. Dept. of Linguistics, Uppsala University.
- Cunningham-Andersson, U. and Engstrand, O. (1989). Perceived strength and identity of foreign accent in Swedish. *Phonetica* **46** (4), 138-154.
- Cunningham-Andersson, U. (2003). Temporal indicators of language dominance in bilingual children. Proceedings from Fonetik 2003, *Phonum* 9, 77-80, Umeå University.
- Cunningham, U. (2004). Language dominance in early and late bilinguals. Swedish Association for Applied Linguistics. Höstsymposium (2004). *Språk på tvärs : rapport från ASLA:s höstsymposium, Södertörn, 11-12 november 2004*. Uppsala: Association suédoise de linguistique appliquée (ASLA).
- Eriksson, A. and Wretling, P. (1997). How flexible is the human voice? – A case study of mimicry. In *Proceedings of EUROSPEECH '97*, Vol. 2, 1043–1046.
- Flege, J.E. and Hammond R. (1982) Mimicry of non-distinctive phonetic differences between language varieties. *Studies in Second Language Acquisition*, **5**, 1-18.
- Flege, J.E. (1984) The detection of French accent by American listeners. *JASA* **76**, 692-70759-70.
- Gimson, A.C. (1974) *An Introduction to the Pronunciation of English* 2nd ed. London: Edward Arnold.
- Jassem, W. (2003) Polish. *Journal of the International Phonetic Association* **33**(1), 103-107.
- Jongman, A. and Wade, T. (2007). Acoustic variability and perceptual training. In: Bohn, Ocke-Schwen (ed.) *Language Experience in Second*

- Language Speech Learning: In Honor of James Emil Flege*, Amsterdam, NLD: John Benjamins Publishing Company.
- McAllister, R., Flege, J.E., Piske, T. (2002) The influence of L1 on the acquisition of Swedish quantity by native speakers of Spanish, English and Estonian. *Journal of Phonetics* **30**, 229-258.
- Oh, Y.R., Yoon, J.S. and Kim, H.K. (2007) Acoustic model adaptation based on pronunciation variability analysis for non-native speech recognition. *Speech Communication* **49**, 59-70.
- Trudgill, P. (1974) Sex, covert prestige and linguistic change in the urban British English of Norwich. *Language in Society*, **1**, 179-195.
- Wells, J.C., 1982. *Accents of English*. Three volumes. Cambridge: Cambridge University Press.
- Wretling, P. and Eriksson A. (1998), Is articulatory timing speaker specific? - Evidence from imitated voices. In: Peter Branderund and Hartmut Traunmüller (eds.) Swedish Phonetics Conference (1998). *Fonetik 98: Proceedings*. Stockholm University: Dept. of Linguistics. 48 – 51.