

QUALITY, QUANTITY AND INTELLIGIBILITY OF VOWELS IN VIETNAMESE-ACCENTED ENGLISH

UNA CUNNINGHAM

Introduction

This paper attempts to describe and explain some of the phonetic and phonological characteristics of the English spoken by Vietnamese speakers from Hanoi and the particular challenges faced by these speakers in acquiring functional English language skills. Vietnam is in a process of becoming a player in the global community, having joined the World Trade Organisation in 2007 and becoming a non-permanent member of the United Nations Security Council in 2008. Although other languages such as Chinese, French and Korean are also used for specific international relationships, the need for English as a language of international communication has never been greater. While many Vietnamese speakers have written and oral comprehension skills which are fairly unproblematic, and many have good written proficiency, their speech is often not immediately intelligible to anyone not accustomed to the English spoken by native speakers of Vietnamese.

The reasons for this are many and complex and have a good deal to do with differences between the sound system of Vietnamese and that of English. Another major factor is presumably the limited opportunities available to the residents of Hanoi to hear English spoken by other than Vietnamese speakers. Another is the lack of access to or familiarity with supplementary teaching materials. Solving these problems is beyond the scope of this paper, but it seems clear that teachers who have difficulty producing English speech which is intelligible to non-Vietnamese listeners will not be able to identify intelligibility problems in their students' production. As models for their own students, these teachers will be compounding their students' intelligibility problems. The result of this is in any case a characteristic Vietnamese accent of English. A number of studies of Vietnamese-accented English do exist, from a contrastive analysis (e.g. Nguyen 1970) to a cross-linguistic analysis (e.g. Tang 2007), but instrumental studies of this accent of English are uncommon.

Kachru (1985) presented a model of the spread of English in the world as three concentric circles: the inner circle where English is spoken as a native language by much of the population, the outer circle where English is spoken as a second language with some kind of official status, and the expanding circle where English

is learned and spoken as a foreign language by those who are not native speakers of English. Other studies of the English spoken by Vietnamese speakers have been set in inner circle situations where English is the community language, in the U.S. (e.g. Tang 2007) or Australia (e.g. Ingram & Nguyen 2007, Nguyen 1970). In this study the focus is on Vietnamese speakers with a monolingual upbringing who have learned English in a classroom setting and who already are or are planning to be teachers of English in the same setting. So the English we are talking about here is very much English as a foreign language, in Kachru's expanding circle. The students gain familiarity at university level with the phonology of RP, but generally do not get a lot of opportunity to speak with non-Vietnamese speakers. Vietnamese students at all levels are, however, often encouraged by their teachers to speak as much as possible to "foreigners" (by which they presumably mean people who do not appear to be Vietnamese), who are of course as likely to be speakers of French, German or Swedish as to be native speakers of English. The use of English as a language of international communication with other non-native speakers is not, however, an obvious or explicit target, and the vowels of RP appear to be the model and target of choice for many teachers.

In a preliminary, informal experiment conducted during a pronunciation class held in Hanoi for young university teachers of English, it was found that the Vietnamese-speaking participants pronounced the English words *bead*, *beat*, *bid*, *bit* in citation form in such a way that the non-Vietnamese-speaking teacher (the author) could not tell which word was being uttered at better than chance although the rest of the class (all Vietnamese speakers) were apparently able to identify the word being pronounced with some accuracy. There is some evidence in the literature that intelligibility is better for those who share the speakers' L1 than for speakers of another L1 (e.g. Derwing, Rossiter and Munro 2002, Jenkins 2002, Smith and Bisazza 1982) while Yule, Wetzel and Kennedy (1990) found that non-native speakers were able to understand their own speech better than that of others who share their L1. On the other hand there are studies which indicate that there is little difference between ratings of non-native speech from native and non-native judges, such as Munro, Derwing and Morton (2006) who found that Japanese-speaking listeners understood Japanese-accented English better than did native English listeners, but not better than Chinese speaking listeners, suggesting that native speakers are uncommonly bad at understanding accented speech. Bradlow and Bent (2008) found that English-speaking judges could be trained to better understand Chinese-accented English, and Major, Fitzmaurice, Bunta and Balasubramanian (2002) found only small and variable advantages to listeners who share the speakers' L1, such that native speakers of Spanish achieved significantly higher results when listening to Spanish-accented speech, but native speakers of Chinese found it significantly more difficult to understand speakers who shared their L1.

The most immediately striking characteristic of the kind of Vietnamese accent of English that I am referring to here is the elision of consonants, in particular final consonants and consonant clusters in the syllable coda. Another feature is that final stops may not be released. This is certainly a major factor in the perceived unintelligibility of Vietnamese accents of English, yet there are many other characteristics of these accents. The *bead-beat-bid-bit* confusion experienced by the non-Vietnamese-speaking listener mentioned above may have more to do with vowel duration and spectral quality than with the articulation (or non-articulation) of the consonant in the coda of the syllable. It is well documented that the primary cue to postvocalic voicing in standard accents of English is vowel duration and that spectral cues are more salient than duration in the distinction between the vowels of *bit* and *beat* (e.g. Flege 1997). This paper is an attempt to describe the vowel systems in operation in several groups of speakers in Hanoi.

Materials and methods

Informants

There are three groups of Vietnamese-speaking informants in the production part of this study, all of them involved as staff or students at a university in Hanoi. Group 1 is made up of seven female university administrators and academic staff aged from 25-45 who have not studied English as a major at university. Group 2 is made up of six female university teachers of English aged 23-30. These were the same individuals who took the pronunciation class mentioned above as a preliminary study. Group 3 is made up of three female English major undergraduate students aged 20-21. A comparison is made with a group of seven Swedish-speaking 16-year old females. Two Vietnamese speakers (females between 23 and 30) who arrived in Sweden from Hanoi just one month before they participated in the study, and six native speakers of English took part in the perception part of this study.

Material

There are two sets of material involved in this study. Set 1 is a text and a list of 44 words reproduced here in Appendix A. This material has earlier been recorded by other speaker groups, including the Swedish-speakers mentioned above. The words of the wordlist occur in the text as well, and 36 of them are chosen to represent nine of the 24 word classes described for native speakers of English by Wells

(1982). These nine word classes are fairly monophthongal in most inner circle varieties of English and do not involve postvocalic /r/. Words in each class are expected to be represented by similar sounding vowels with no phonemic distinctions being made within a word class when spoken by native speakers of English. Different accents of English are then expected to have different combinations of distinctions made between word classes. Wells does describe some second language varieties of English, such as Singapore English and Filipino English, but the application of his model to the phonology of non-monolingual English speech may well be difficult, given the variability that is said to be characteristic of non-native speech (cf e.g. Jenkins 2000, Cunningham 2008). However the model does make it possible to refer to words without committing to a phonemic analysis, which is an advantage in the case of non-native speech.

Table x-1 Words and word classes

Word class	Words in the material
DRESS	<i>friends, very</i>
FLEECE	<i>believe, feel, green, leaves, see, sheep, trees</i>
FOOT	<i>could, pull, would</i>
GOOSE	<i>choose, pool, room, school, through</i>
KIT	<i>grin, quickly, ship, still, think, this, window</i>
LOT	<i>because, longer</i>
STRUT	<i>become, comfort, country, govern, run, shut</i>
THOUGHT	<i>small, thought</i>
TRAP	<i>man, unhappy</i>

A more extensive set of material was elicited from group 3 and an RP-speaking control. The second set of stimuli actually includes the first set, but also includes a battery of words embedded in carrier phrases to make it possible to study temporal effects of the elicited speech, such as the duration of vowels and postvocalic consonants in a variety of conditions such as with different vowels, different postvocalic consonant voicing, and mid vs phrase-final position of the test word. Examples of set two sentences are shown in Appendix A.

Major (2001:63) points out that “[in] L1 and L2 acquisition, learners generally approximate the target with greater accuracy with increasing formality”. He goes on to suggest that wordlists, with their focus on the form rather than content will elicit the most accurate pronunciation. So, without necessarily adopting the “L2 user as a deficient native speaker” view of accented speech condemned in Cook (2002: 63) and elsewhere which is inherent in Major’s text, his suggestion that the formal wordlist and text material will in any case ensure that the informants are able to pay maximal attention to their pronunciation seems intuitively attractive.

Method

The informants in group 1 and group 2 were recorded directly into the computer using a headset and WaveSurfer (Sjölander & Beskow 2000). They read the stimuli from paper, reading the text once and the word list twice. The informants in group 3 were recorded using a Zoom H4 digital recorder and the stimuli were presented using Microsoft PowerPoint. It proved to be difficult to arrange optimal sound recording conditions at the university in Hanoi, and some items were impossible to analyse and these speakers were excluded from this study.

Measurements were made of the material using Praat (Boersma & Weenink 2008). F1 and F2 values were measured for the vowels in the words in set one listed in table 1. The formant measurements were made as average values over 50 ms of steady vowel quality (where possible). For set two, durational measurements were made of the vowels and postvocalic consonants and formant measurements were made of F1 and F2 for the KIT and FLEECE vowels in the two stimuli subsets *bead*, *beat*, *bid*, *bit* and *seed*, *seat*, *Sid*, *sit*.

Results and Discussion

Vowel space

It is particularly interesting to see which vowel quality distinctions the speakers maintain. This may indicate phonemic oppositions they are observing. Within-category variation may suggest that the speakers' category boundaries do not quite coincide with Wells' word categories.

Fig. x-1 shows the average F1 and F2 values in Hz for the word categories for the seven speakers in Group 1. These speakers have not studied English as a major at university level. Each word was uttered twice by each speaker. The data is shown with linear axes. Compare this with the corresponding material in Fig. x-2 for the speakers in Group 2, who have graduated from an English major programme and who are university teachers of English. They have had explicit pronunciation teaching and training including a course given by the author shortly before the recordings were made. Thus Group 1 has lower overall English language proficiency than Group 2.

Fig.x-1 Group 1: average values for F1 vs F2

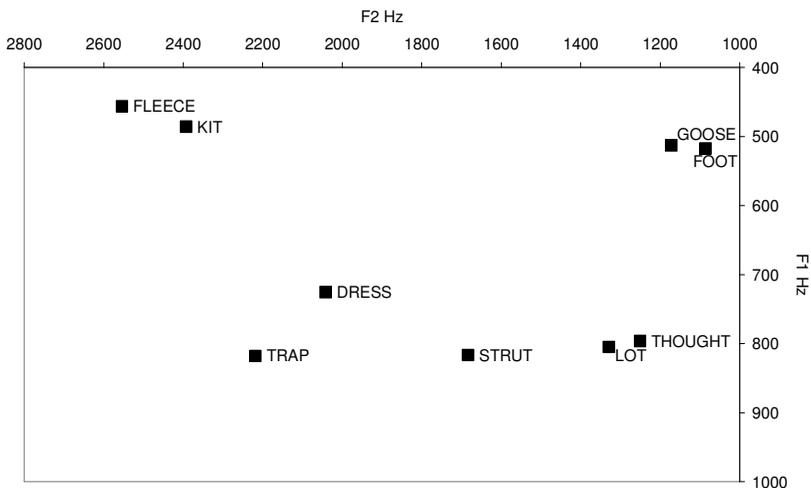
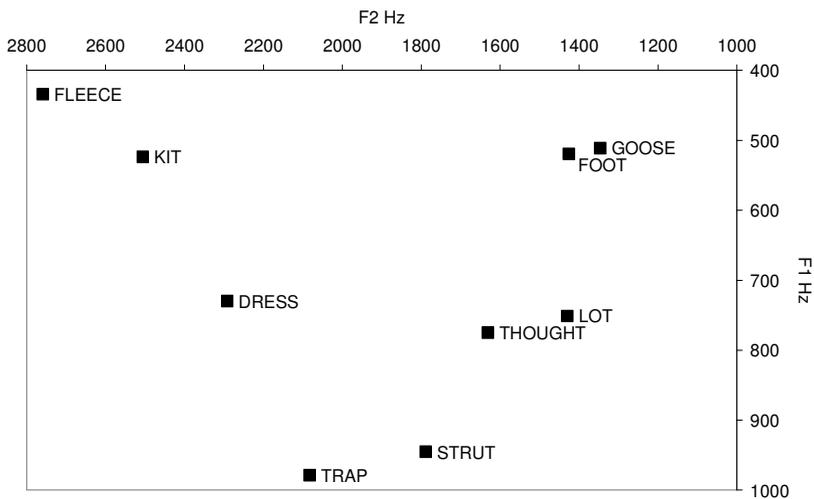


Fig. x-2 Group 2: average values for F1 vs F2



There is considerable difference between the two figures. Group 1 have little difference between the average vowel qualities associated with FLEECE and KIT, GOOSE and FOOT, and LOT and THOUGHT. Group 2, on the other hand, seem to have a clearer separation at least between the average formant frequencies for

the vowels in the words representing FLEECE and KIT. There does seem to be very little difference in quality between GOOSE and FOOT and between LOT and THOUGHT at least for group 1. But these word categories are not distinguished by all groups of monolingual English speakers. FOOT and GOOSE merge in Scottish English and Northern Irish English (Wells 1982, Cunningham 2008). LOT and THOUGHT are not distinguished in many American and other accents (Wells 1982). The semantic load for at least the opposition between GOOSE and FOOT is singularly low. In fact, there appear to be only three minimal pairs that do not involve morpheme boundaries (*Luke-look, pool-pull, fool-full*). So a failure to distinguish these pairs is less problematic for intelligibility than failure to observe the KIT-FLEECE distinction. Jenkins (2002:97) identifies this vowel pair as one of the more important targets for learners to master.

FLEECE vs KIT

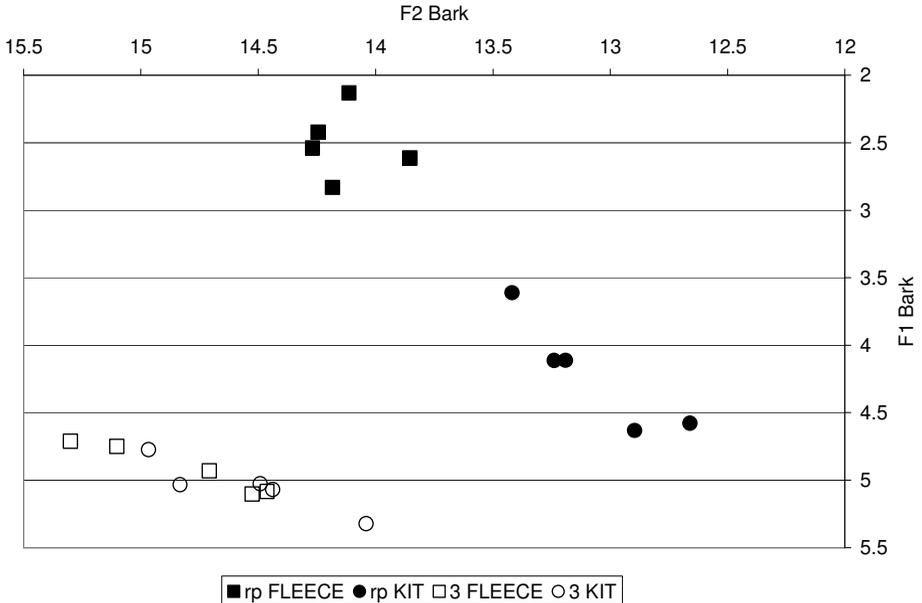
So it seems that the separation of the FLEECE and KIT categories is not quite clear in the English spoken in Hanoi. In other native varieties this distinction is often carried by both a vowel quality difference and a vowel quantity difference, such that in the major model varieties RP and GenAm, the FLEECE vowel /i:/ is usually longer, higher and more front than the KIT vowel /ɪ/. In acoustic terms we can say that the FLEECE vowel usually has lower F1 and Higher F2 and greater duration than the KIT vowel. The three speakers in group 3, who are expected to have comparable proficiency to the group 2 speakers, but who did not take part in the pronunciation class offered by the author, recorded a larger set of data where the FLEECE and KIT vowels occurred in a number of different contexts. This facilitates an in depth study of the quality and quantity characteristics of the vowels. These characteristics will be studied here by measurements of the acoustic signal.

The task here is to establish whether the three group 3 speakers are in fact reliably distinguishing between the vowel quality in KIT and FLEECE vowels. In Flege (1987) the concept of equivalence classification is developed concerning sounds which are identical, similar or new in the L2. The question here is whether the KIT vowel is being classified along with the FLEECE vowel as similar to the Vietnamese high front unrounded vowel or whether one of the vowels is being distinguished as a new vowel.

A comparison can be made between their pronunciation of FLEECE words like *beat, bead, seat, seed, sheep* and KIT words such as *bit, bid, sit, Sid, ship*. Fig. x-3 shows these vowels as F1 plotted against F2 in Bark to better reflect the auditory relationship between the values. The words were uttered three to five times by each speaker in the carrier phrase “*I’m saying _____ again*” and the average values for

F1 and F2 are plotted. The figure also shows the average values from the same words uttered three to five times each by an RP speaker, for the purpose of comparison.

Fig. x-3 Group 3 and an RP speaker: F1 vs F2 values for the vowels of KIT and FLEECE words

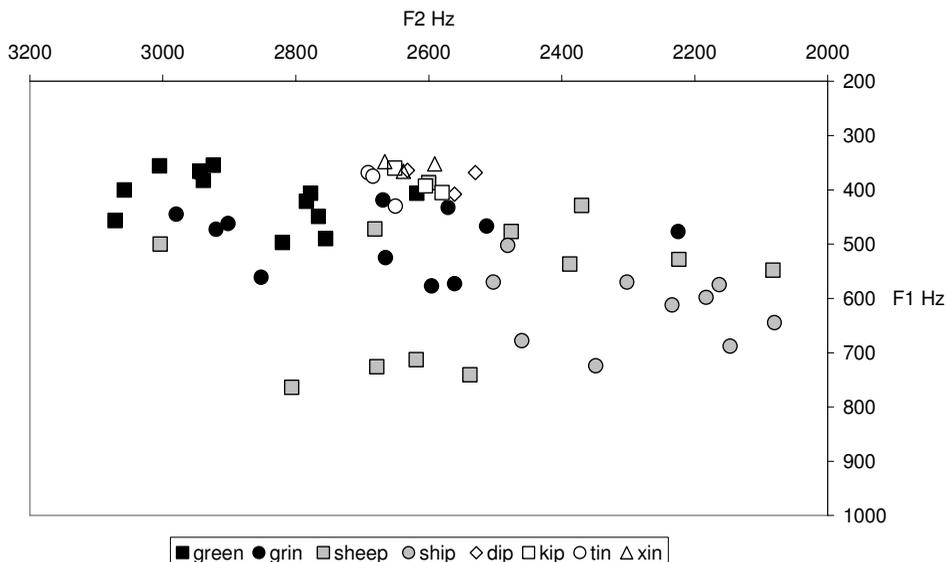


Where the RP speaker has a clear separation in the F1-F2 space between the FLEECE vowels and the KIT vowels, the Group 3 speakers pronounce the vowels without distinguishing F1 or F2 in the same way. There is considerable overlap between the categories, but a 1-tailed t-test shows that there is in fact a significant difference in F1 and F2 for the FLEECE vs KIT vowels ($p(t) < 0.01$) for both tests). The KIT vowels have lower average F2 and higher average F1 than the FLEECE vowels. So the speakers are clearly aiming at two different target values, even though there is overlap between the formant frequencies of the vowels produced.

There may be substrate effects from Vietnamese phonology here. The average values shown in the above figures mask a good deal of overlap between the word categories. Within the FLEECE and KIT categories different words appear to have some variation. Fig. x-4 shows the frequencies of F1 vs F2 for the pairs of words *sheep-ship*

between the F1 or F2 values for *green* and *grin* and between *sheep* and *ship* are not shown by t-tests to be significant, but there are highly significant differences in two-tailed t-tests between the F1 and F2 values for *green/grin* vs *sheep/ship* ($p(t) < 0.0001$ for both F1 and F2).

Fig. x-4 Group 2: F1 vs F2 values for the vowels of *grin*, *green*, *sheep* and *ship* and *xin*, *kip*, *tin*, *dip* .



A number of explanations are possible for this phenomenon. We could explain it as what e.g. Major (2001) calls universal factors, or as coarticulation, or as an artefact of the articulatory system where the F2 can be influenced by the physical properties of the vocal tract during vowel articulation after the settings required to produce the prevocalic consonant or in anticipation of the articulation of the following consonant. Or it could be an effect of transfer from Vietnamese.

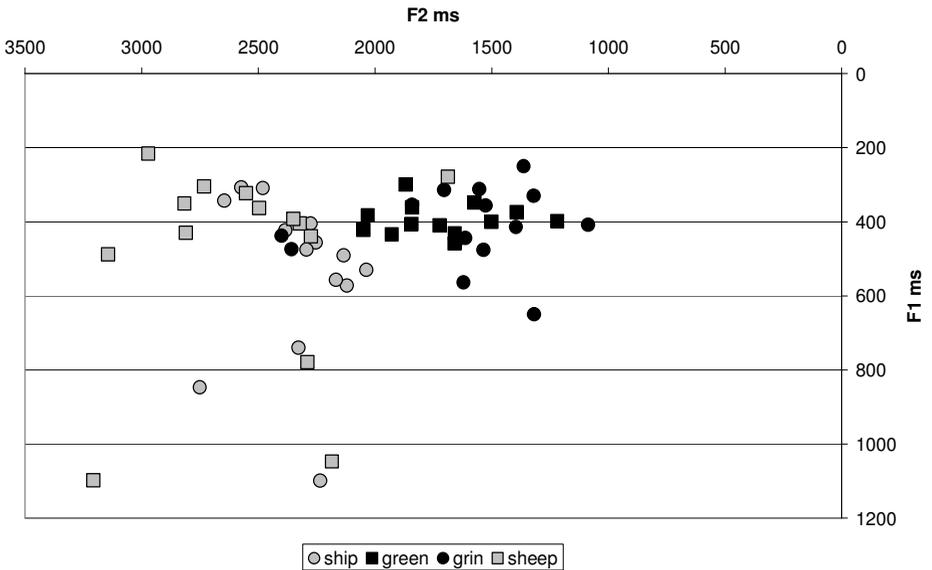
Now, according to Nguyen (1970:131) and Thompson (1987:30), the precise quality of the high front unrounded vowel in the Vietnamese language is said to exhibit allophonic variation. Thompson claims that when this vowel occurs before certain sounds including /p/ it will be higher (“upper high front”) than if it occurs before certain other sounds including /n/ (“lower high front”). Nguyen, on the other hand, suggests that when the vowel occurs before /p/ in such words it will be more front and lower than before /n/ (Nguyen 1970:131). Neither offer spectrographic

support for their claims, making them difficult to compare. So it seems that there is an audible difference in vowel quality depending on the postvocalic consonant, but it is not clear which direction that difference takes.

In order to resolve this issue, recordings were made of a female Vietnamese speaker from Hanoi (VN2) uttering the four words *díp* (“heavy (of eyes)”), *tin* (“news”), *kip* (“be urgent”), *xin* (“to ask for”) three times each in citation form. The results are plotted in Fig. x-4, alongside those of group 2 uttering the English words *sheep*, *ship*, *green* and *grin*. There is an effect in the Vietnamese words such that the vowels followed by a bilabial have lower F2 ($p(t) < 0.01$) than the vowels followed by dentals/alveolars, but the difference in the Vietnamese words is clearly very small compared to the extensive variability demonstrated by group 2 for the English words.

If this were to be some kind of universal phenomenon, it would be found in other accents of English as well. By way of comparison, to establish if this is a general phenomenon or specific to Vietnamese speakers, consider fig. x-5, which shows the same words uttered by seven 16-year old female Swedish speakers.

Fig. x-5 Swedish speakers: F1 vs F2 values for the vowels of *grin*, *green*, *sheep* and *ship*



In this case too a t-test shows that there is no significant difference between the F1 or F2 values for *sheep* and *ship* or for *green* and *grin*. These words are not distinguished from each other spectrally by these speakers either. Like the

Vietnamese speakers, however, this figure shows clear separation between the formant frequencies of F1 and F2 in *sheep/ship* and *green/grin* for these Swedish speakers. However in this case the vowels followed by /p/ have higher F2 than those followed by /n/. Or rather, the vowels followed by /p/ have similar F1 and F2 frequencies for both Swedish and Vietnamese speakers while the vowels followed by /n/ have higher F2 for the Vietnamese speakers. This suggests that the Vietnamese speakers are subject to an effect that the Swedish speakers are not affected by, indicating transfer rather than universals in the terms used by Major (2001:65), or perhaps more neutrally we can say that there is a substrate effect from Vietnamese in accordance with the findings of Thompson (1987) and Nguyen (1970) for Vietnamese. So it seems likely that the speakers in both group 1 and group 2 might be influenced by Vietnamese phonological processes in their pronunciation of this vowel.

It has been shown that the less proficient group 1 speakers have no significant difference between the F1 and F2 values for KIT vowels and FLEECE vowels while the more proficient group 2 and three speakers have significant differences between the F1 and F2 of the vowels, but have extensive overlap between the categories. Do they, like monolingual English speakers use duration as an additional cue to the distinction between KIT and FLEECE vowels?

Fig. x-6 Group 3 and RP speaker: Vowel durations for the vowels *bead*, *beat*, *bid*, *bit* in non-phrase final position

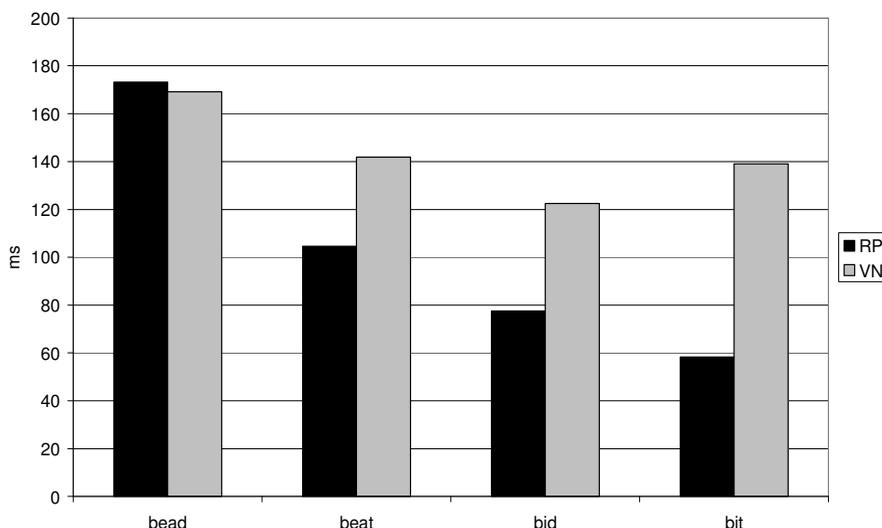


Fig. x-6 shows the average durations of the vowels and of the postvocalic stops for 3-5 tokens of each of the words *bead*, *beat*, *bid* and *bit* as uttered in non final position (in the carrier phrase *I'm saying ____ again*) by the three speakers in group 3 and an RP speaker. The Vietnamese speakers produce similar vowel duration in *beat* and *bit*, while the vowel in *bead* is much longer than that in *bid*. This suggests that the duration of the vowel is not systematically used as a cue to its identity by the Vietnamese speakers. The RP speaker, as expected, produces longer vowels in *bead* and *beat* than in *bid* and *bit* respectively. Notice also in Fig. x-6 that there is little pre-fortis clipping apparent for the Vietnamese speakers. In fact the vowel in *bit* is conspicuously longer than the vowel in *bid* for the Vietnamese speakers. The RP speaker demonstrates this both for the vowel in *beat* compared to *bead* and *bit* compared to *bid*, but does not go to the lengths suggested by Gimson (through Cruttenden 2008:95), who suggests that the vowel of *beat* be about half as long as the vowel of *bead*.

Perception

This study was designed to examine the intelligibility difficulties experienced by the English-speaking author when listening to Vietnamese speakers speaking English, but apparently less so by the Vietnamese-speaking classmates of speakers in the pronunciation class mentioned at the beginning of this paper. In order to test this anecdotal finding, two female speakers of Vietnamese from Hanoi participated (VN1 and VN2). They had good English language proficiency and could be understood fairly easily in conversation. Each read the battery of stimulus material, and three tokens of each of the four utterances *I'm saying bead/beat/bid/bit again* were presented to a panel of six native English-speaking listeners, and also to the two Vietnamese speakers VN1 and VN2. The same material as spoken by the English-speaking author was presented to the two Vietnamese speakers and to two of the native English-speaking listeners as a control.

Fig. x-7 Perception of English and Vietnamese speaking speakers by English and Vietnamese speaking listeners

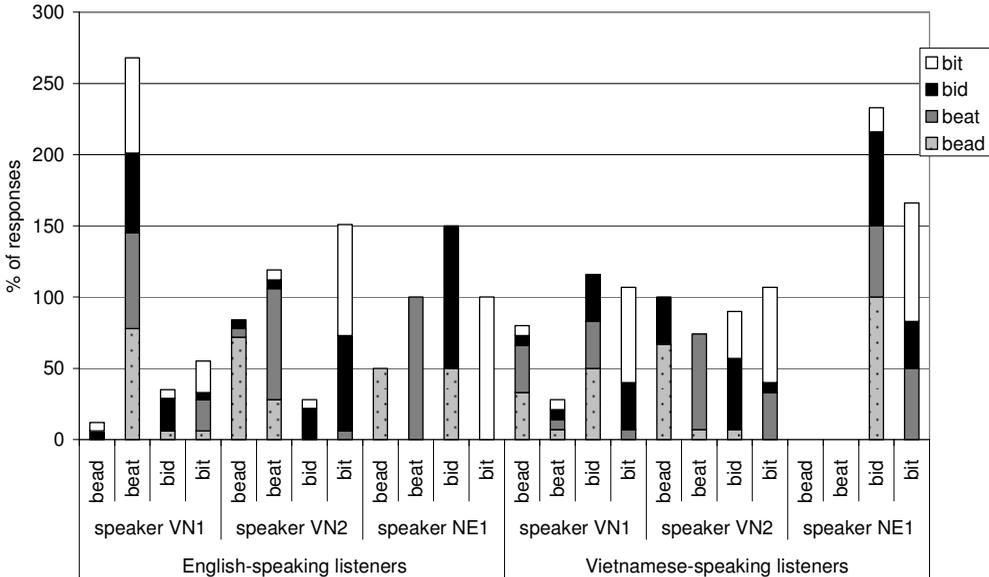


Fig. x-7 shows the perception of the Vietnamese speakers' English with the native English-speaking control as perceived by English-speaking and Vietnamese-speaking listeners. The native English-speaking listeners perceived VN1's vowel correctly in 55% of cases and her consonant correctly in 58%, which is little better than chance. In fact, most of what she said was heard as *beat*. The Vietnamese-speaking listeners (VN1 and VN2 themselves) did slightly better and heard the vowel correctly (as intended by the speaker) in 58% of cases and the consonant in 67% of cases. Interestingly, unlike the native speakers of English, they hardly ever heard *beat* in what VN1 said. This is a clear indication that the cues being listened for by the Vietnamese speakers are not the same as those being transmitted or expected by the native listeners. For speaker VN2, the native English speakers heard the vowel correctly in 89% of cases and the consonant in 69%. In fact, the native English listeners on average heard the intended vowel rather better than VN2 herself did, although they often mistook *bid* for *bit*. This is in line with the findings reported by Major, Fitzmaurice, Bunta and Fujieda (2002) and fails to support the findings of e.g. Derwing, Rossiter and Munro (2002) and Jenkins (2002). The average results for the two Vietnamese-speaking listeners listening to VN2's utterances were 79% correct vowel and 83% correct consonant.

The control case of the native speaker stimuli showed that the native speaker was much more intelligible to the native listeners, although one of the two listeners, an RP speaker, misheard the intended *bead* as *bid* in all three tokens. The Vietnamese-speaking listeners found it impossible to interpret any of the native speaker stimuli as containing instances of the FLEECE vowel. Thus they were 50% accurate for the vowel and 75% accurate in the case of the consonant.

Conclusion

The Vietnamese speakers studied had varying degrees of training and proficiency in English. They all have some degree of difficulty producing speech that is intelligible. They tend to rely heavily on contextual cues to get their message across. This study has shown some of the reasons why intelligibility is a problem for these speakers.

A vowel plot (figs. x-1 and x-2) showed that there is an apparent merger between the word classes denoted by Wells (1982) as KIT and FLEECE, GOOSE and FOOT and LOT and THOUGHT respectively. While the GOOSE-FOOT and LOT-THOUGHT mergers carry a low semantic load, and are merged in some native varieties of English, the KIT-FLEECE is one which is generally held as important to uphold for purposes of intelligibility.

On closer study it became apparent that although the most proficient speakers, those in group 2 and 3, were in fact producing a significant difference between the F1 and F2 frequencies for the KIT (*bit, bid, sit, Sid, ship*) and FLEECE (*beat, bead, seat, seed, sheep*) vowels, there was considerable overlap between the categories, as can be seen in Fig. x-3 for group 3 compared to an RP speaker (who has a clear distinction between KIT and FLEECE in the F1-F2 vowel space).

In an attempt to explain this overlap, and the variation in the precise quality of the vowel in different words, the F1 and F2 of the vowels of the word pairs *grin, green* and *ship, sheep* were plotted in Fig. x-4, alongside the F1 and F2 of vowels in four Vietnamese words also ending in the high front unrounded vowel /i/ followed by /p/ or /n/. According to Nguyen (1970) and Thompson (1987), there is a difference in quality of this vowel conditioned by the place of articulation of the following consonant. This was indeed found in the Vietnamese speech produced, such that the /i/ was realised with a higher F2 if followed by /n/ than if followed by /p/. In the English words a similar effect was found such that the vowels of the words *sheep* and *ship* had significantly lower F2 than the vowels of the words *green* and *grin*. The difference was in the range of variation. While there was about 200 Hz between the highest and lowest F2 value for the Vietnamese material, there was about 1000 Hz between the highest and lowest values for the English material.

To examine the possibility that this might be a universal coarticulatory effect, similar material from English spoken by young female Swedish speakers was compared. In this material a similar overlap between the positions of KIT words and FLEECE words in F1-F2 space was found, and an F2 difference between *sheep* and *ship* where the vowel is followed by /p/ and *green* and *grin* where the vowel is followed by /n/ was also found. The interesting thing here is that the difference was in the opposite direction. The F2 in words where /p/ is the final consonant was higher than that where /n/ was the final consonant. So it is clear that this effect is language specific; not a universal and likely due to transfer from the substrate language. A possibility here might be that the F2 difference is an artefact of differences in lip-rounding between the Vietnamese and Swedish speakers in the production of the prevocalic /r/ and /ʀ/

Durational differences in vowels are heavily used by monolingual English speakers and listeners as secondary cues to vowel identity and primary cues to postvocalic voicing. It was found that even the relatively proficient speakers in group 3 mastered neither of these systematic distinctions in their production.

As regards perception, it was found that the native English listeners had severe problems identifying which of the words *bead*, *beat*, *bid*, *bit* the Vietnamese-speaking informants had intended when the words were presented in carrier phrases which removed contextual cues. The Vietnamese speakers were only a little better at hearing what they had intended themselves. The English speakers were generally able to hear native English without difficulty except for an RP speaker who consistently heard a Northern Irish *bead* as *bid*. The Vietnamese speakers heard the Northern Irish *bead* and *beat* as *bid* or *bit* with apparently arbitrary perception of the intended consonant. Both the Vietnamese and the RP listener may well have done better with RP stimuli. This can be compared with a study by Matsuura, Chiba and Fujieda (1999) who found that Japanese listeners found Irish English intelligible but may prefer more familiar varieties. This leads us to the very interesting matter of where the line goes between a need to improve speaker intelligibility on the one hand, and a need to improve the flexibility of native and non-native listeners on the other. The speakers in this study could not reliably understand their own speech in the case of these distinctions that are more or less merged in their speech. Unless they learn to make these distinctions they are bound to rely on contextual cues to get their message across even to other speakers of Vietnamese. At the same time, as long as native listeners are unable to correctly perceive distinctions that are in fact being made but that do not employ the salient cues they are expecting to hear (cf Derwing, Rossiter, and Munro 2002), there will be problems.

References

- Boersma, P. & Weenink, D. 2009. *Praat: doing phonetics by computer* (Version 5.1.01) [Computer program]. Accessed on 28 February 2009, from <http://www.praat.org/>.
- Bradlow, A.R. and Bent, T. 2008. Perceptual adaptation to non-native speech. *Cognition* 106, 707-729.
- Cook, V. J. 2002. Background to the L2 User. In: *Second Language Acquisition, 1: Portraits of the L2 User*, ed. V. J. Cook. Clevedon: Multilingual Matters Limited 1-28.
- Cruttendon, A. 2008. *Gimson's Pronunciation of English* 7th ed. London: Hodder Education.
- Cunningham, U. 2008. Acoustic variability in the production of English Vowels by Native and Non-Native Speakers. In *Variability in Accents of English: A non-native perspective*, ed. E. Waniek-Klimczak. Cambridge: Cambridge Scholar Press.
- Cunningham, U. 2008. Vowels in rural southwest Tyrone. *Proceedings, FONETIK 2008*, Department of Linguistics, University of Gothenburg. Accessed on 28 February 2009 at http://www.ling.gu.se/konferenser/fonetik2008/papers/fon08_Cunningham.pdf.
- Derwing, T. M., Rossiter, M. J. and Munro, M. J. 2002. Teaching native speakers to listen to foreign-accented speech. *Journal of Multilingual and Multicultural Development* 23, (4) 245-259.
- Flege, J. E. 1987. The production of "new" and "similar" phones in a foreign language: evidence for the effect of equivalence classification. *Journal of Phonetics* 15, 47-65.
- Flege, J. E., Bohn, O-S., Jang, S. 1997. Effects of experience on non-native speakers' production and perception of English vowels. *Journal of Phonetics* 25, 437-470.
- Ingram J. C. L. and Nguyen T. T. A. 2007. *Vietnamese accented English: Foreign accent and intelligibility judgement by listeners of different language backgrounds*. Unpublished ms, University of Queensland. Accessed on 2 August 2008 after personal communication at: http://www.ads.edu.vn/ads/uploads/file/ADS%202009/TESOL_Conference/Ingram%20&%20Nguyen-Vietnamese%20accented%20English%20Foreign%20Accent%20and%20Intelligibility%20judgement%20by%20listeners%20of%20different%20language%20backgrounds.pdf.
- Jenkins, J. 2000. *The Phonology of English as an International Language : New Models, New Norms, New Goals*. Oxford: Oxford Univ. Press.

- Jenkins, J. 2002. A sociolinguistically based, empirically researched pronunciation syllabus for English as an International Language. *Applied Linguistics* 23 (1) 83-103.
- Kachru, B. B. 1985. Standards, codification and sociolinguistic realism: The English language in the outer circle. In *English in the World: Teaching and Learning the Language and Literatures*, eds. R. Quirk and H. Widdowson. Cambridge: Cambridge University Press and The British Council, 11-30.
- Major, R. C. 2001. *Foreign Accent: The Ontogeny and Phylogeny of Second Language Phonology*. Mahwah, N.J.: Lawrence Erlbaum Associates.
- Major, R. C., Fitzmaurice, S. Bunta, F. and Balasubramanian, C. 2002. The effects of non-native accents on listening comprehension: Implications for ESL assessment. *TESOL Quarterly*, 36, 173-190.
- Matsuura, H., Chiba, R. and Fujieda, M. 1999. Intelligibility and comprehensibility of American and Irish Englishes in Japan. *World Englishes*, 18(1), 49-62.
- Munro, M. J., Derwing, T. M., and Morton, S. L. 2006. The mutual intelligibility of foreign accents. *Studies in Second Language Acquisition*, 28, 111-131.
- Nguyen, D. L. 1970. A contrastive phonological analysis of English and Vietnamese. *Pacific Linguistics Series*, no. 8. Canberra: Australian National University.
- Riney, T. J. 1988. *The Interlanguage Phonology of Vietnamese English*. Unpublished Ph.D. dissertation, Georgetown University.
- Sjölander, K. and Beskow, J. 2000. Wavesurfer - an open source speech tool. Paper presented at the ICSLP 2000 conference in Beijing, China. Accessed on 20 October 2008 at: http://www.speech.kth.se/wavesurfer/wsurf_icslp00.pdf.
- Smith, L.E. and Bisazza, J.A. 1982. The comprehensibility of three varieties of English for college students in seven countries. *Language Learning* 32, 259-69.
- Tang, G. M. 2007. Cross-linguistic analysis of Vietnamese and English with implications for Vietnamese language acquisition and maintenance in the United States. *Journal of Southeast Asian American Education and Advancement*, 2. Accessed on 1 August at: <http://jsaeea.coehd.utsa.edu/index.php/JSAAEA/issue/view/5>.
- Thompson, L. C. 1987. *A Vietnamese Reference Grammar*. Honolulu: Univ. of Hawaii Press.
- Wells, J. C. 1982. *Accents of English Vol. I-III*. Cambridge: Cambridge Univ. Press.
- Yule, G., Wetzel, S. and Kennedy, L. 1990. Listening perception accuracy of ESL learners as a variable function of speaker L1. *TESOL Quarterly* 24, 519-23.

Appendix A Stimuli material

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.

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

Examples of stimuli in Stimulus set 2

I'm saying bead / beat / bid / bit again

I'm saying seed / seat / Sid / sit again

I'm saying edge / etch again

I'm saying eyes / ice again

I'm saying figs / fix again

I'm saying ship / sheep again

Now I'm saying bead / beat / bid / bit