

Teaching English Intonation with a Visual Display of Fundamental Frequency

Richard Stibbard

<http://www.hkbu.edu.hk/~stibbard/>

[Hong Kong Baptist University](http://www.hkbu.edu.hk/~stibbard/)

Introduction

The teaching of pronunciation in a communicative approach

In a reaction against the theory and practice prevalent from the 1960s until the early 1980s, good pronunciation skills are now increasingly being seen as important in a communicative approach to teaching English as a Foreign/Second Language. Indeed, Morley argues that pronunciation belongs at the very core of a communicative approach to language teaching, writing: "[i]ntelligible pronunciation is an essential component of communicative competence" [Morley](#) (1991:488). She argues for four "reasonable and desirable" learner goals, namely: "functional intelligibility, functional communicability, increased self-confidence and speech monitoring abilities and speech modification strategies for use beyond the classroom". ([Morley](#), 500).

In the context of this study, involving university students in Hong Kong, Morley's assertion that students need "instruction that will give them communicative empowerment - effective language use that will help them not just to survive, but to *succeed*" ([Morley](#), 489; my italics) could not be more true. Both to succeed on a personal level and to contribute to the continued success of Hong Kong on the world stage, many of our students will need a standard of spoken English high enough to feel at ease in demanding business or academic situations involving participants from around the world.

In order to equip them for these tasks, the teacher of English pronunciation in Hong Kong must address the many facets of pronunciation. It is now widely accepted that pronunciation teaching involves attention not just to the segmental (phonemic) level but to the suprasegmental level as well, which includes those features which span across the phonemes and operate at sentence, discourse or language level.

Brown states that "[w]riters are now convinced of the importance of suprasegmentals in pronunciation" and argues for more attention to intonation in the classroom ([Brown](#) 1995: 172). [Jones and Evans](#) (1995) argue that pronunciation teaching should begin at the level of voice quality and point to the characteristic differences in voice settings in different languages, while Leather, in his influential state of the art paper on the subject notes that "[w]hen L2 has lexical tones [...] the arguments for prosodic training seem all the more compelling" ([Leather](#) 1983: 200). Cantonese learners of English do indeed face great difficulties, both in physiological/phonetic terms, that is, in the physical problem of how to produce authentic sounding intonation, and in phonological/semantic terms, that is, in deciding how to use intonation meaningfully in their speech.

A number of researchers and teachers have used speech analysis programs to teach suprasegmental features of English pronunciation (e.g. [Molholt](#) 1988) and Leather welcomes the use of computer technology to teach suprasegmental features of English pronunciation ([Leather](#), 1983: 211), while Morley refers positively to the possible imaginative uses to which computers might be put ([Morley](#), 1991: 511). Others are at present engaged in further research into the pedagogical applications of visual displays of speech ([Lambacher](#) 1996).

The expansion in higher education in Hong Kong and elsewhere has in recent years led to greater emphasis being placed on autonomous self-access work and enhanced self-monitoring skills on the part of students. This paper reports on a pilot self-access programme in which students used a fundamental frequency analyser to see visual feedback on their intonation patterns. Projects such as this, which place in students' hands tools to help them monitor and improve their own performance are thus valuable in view of the current educational climate.

After discussing the acoustic correlates of the phonological features pitch, intonation and prominence, and giving a brief overview of

the pedagogical model of intonation used, this paper presents examples of the type of feedback generated, illustrating a number of features of student's pronunciation which can be examined, and reports on possible areas of pronunciation this may benefit, before concluding with suggestions regarding the introduction of a CSL-Pitch in a Self-Access Centre.

Pitch, intonation and prominence

Pitch is defined as the relative height of speech sounds as perceived by a listener and is what we are hearing when we refer to a voice being "high" or "low". The varying pitch levels throughout an utterance form what we hear as intonation: the "falling" or "rising" of the voice ([Cruttenden](#), 1986: 4).

Prominence is what we hear when a word "stands out" from those around it, as in for example the prominent word "I" in "I am", a possible answer to "Who's coming?", compared with the prominent word "am" in "I am" answering perhaps "You're not coming, are you?"

The primary physiological cause of both pitch and prominence in speech is the varying rate of vibration of the vocal folds, the acoustic correlate of which is fundamental frequency (F0). The correlation between pitch and fundamental frequency is non-linear: the frequency difference between two tones necessary for listeners to judge that the higher tone is twice as high as the other is much greater at high absolute frequencies than at low. However, as F0 frequencies are relatively low, that is, below 500Hz., pitch can for practical purposes be equated with F0 ([Cruttenden](#) 1986: 4) and indeed the vertical axis of the CSL-Pitch display is labelled "Pitch".

Other factors are involved in prominence and intonation, including duration and loudness, loudness being the (again, non-linear) perceptual correlate of the acoustic feature amplitude. While these factors are relevant, they are generally recognised to be secondary in importance to fundamental frequency ([Cruttenden](#), 1986: 2). Although the CSL-Pitch is capable of displaying energy levels and information on a number of other parameters, these were therefore not the focus of the project and with the occasional exception of duration where this was an important factor, were not discussed.

The pedagogical model

The model of intonation used for this study is that proposed by [Brazil](#) (1985), a concise review of which can be found in [Coulthard](#) (1992). Essentially, the tenets of Discourse Intonation are that the primary function of intonation is interactionally motivated, specific situations influencing the speakers' choices. Of fundamental importance in the model is the notion of "common ground" between the speaker and hearer. Knowledge which is assumed to be common ground will be "referred to"; that which is new information, the focus, will be "proclaimed" as such ([Brazil](#) 1985: 200).

An example will serve to make this clear. Consider this sentence, taken from a government announcement on local television :

"If we want a world that's safe for everyone, we can't do it alone"

In this example, the first clause is spoken with a "referring" fall-rise tone and the second with a "proclaiming" tone, the implication being that the idea of "a world that's safe for everyone" is something familiar to the listener, while the focus, the new idea, is contained in the next half of the sentence, "we can't do it alone".

The following table gives the possible tone choices in interactive discourse as identified by the model and the notation used in this paper to indicate them.

Table 1 The functions and notation used for the tones in direct orientation

Function	Role of speaker	Notation (Brazil, 1985)	Realised as
Referring to common ground	Non-dominant	r	Fall-rise tone
	Dominant	r+	Rising tone

Proclaiming new information	Non-dominant	p	Falling tone
	Dominant	p+	Rise-fall tone

It is important to note that Brazil's tonal categories are arrived at by an initial hypothesis as to the significance of tones in English discourse: "before proceeding to detailed phonetic specification we need to know how many meaningful oppositions there are..." (Brazil 1985: 14). Only secondly does he relate this set of meaningful oppositions to their phonetic realisations, because that is "the only research procedure available" (loc. cit.).

While this approach is a more teachable system than other, attitudinal, approaches to intonation, the fact that the tonal categories proposed are based on both discursal significance and on phonetic realisations leaves the teacher with a problem at the phonetic level. Brazil is indeed careful to emphasise that his descriptions of the phonetic realisations of the tones are not intended to be phonetically precise. Rather, they are intended to be a convenient shorthand by means of which typical realisations can be labelled, whilst recognising the wide range of non-significant variations which can occur within each category. For this reason he uses "prevaricating quotation marks" around all phonetic descriptions (Brazil 1985: 15) and maintains a "mental separation between the meaningful distinctions the speaker makes and the physical events whereby his decisions are manifest". (Brazil 1985: 22). His referring tones are thus characterised at the same time both by their function of referring to the changing common ground in the "interpenetrating biographies" of speaker and hearer as the discourse progresses, and by their typically rising or fall-rising pitch.

Notation

In the figures which follow, the notation of the title follows these conventions. These are adapted from those found in Bradford (1988) and Brazil (1994) for this HTML document as not all browsers show underlining, which was what the students had to follow. On occasions these are in normal orthography with no indication of suprasegmental features. On other occasions, the following conventions apply:

Table 2 The notation used in this HTML document

Item	Notation
Tone unit boundaries	// text //
Prominence	Upper case
Tonic syllable	Underlining
Tones	Codes as in Table 1

The CSL-Pitch analyser

An add-on utility to the Kay CSL Speech Laboratory, the CSL-Pitch is a user-friendly DOS programme with menus at the top of the screen for functions such data capture and playback. The CSL-Pitch produces a split screen display, one half displaying the pitch contour of the model, the other that of the student. (In practice, either view window can be used by the student but, for the sake of clarity, in this paper the upper window is always used to display the student's attempt.) The program allows the student to listen to the voice of a model (stored as a DOS file on the hard disk of the computer) and to record up to fifteen seconds of speech, during which time a pitch contour is simultaneously generated (hence the name "real-time" pitch analyser), providing the student with instant visual feedback on his/her efforts. The student can then listen to his/her voice and to that of the model again, watching the real-time pitch display as it moves across the screen, and can make further attempts at imitating the model.

To begin comparison, the student first uses the mouse to select one of the two view windows, then opens a file containing the sample model's voice. A pitch contour is displayed, which "walks" across the screen in synchrony with the playback. Students can now play this file back as many times as they wish to familiarise themselves with the changing pitch pattern and when ready can begin recording ("capturing") their own voice.

During capture, the CSL-Pitch analyses the F0 levels and produces an instant display which again "walks" across the screen as the

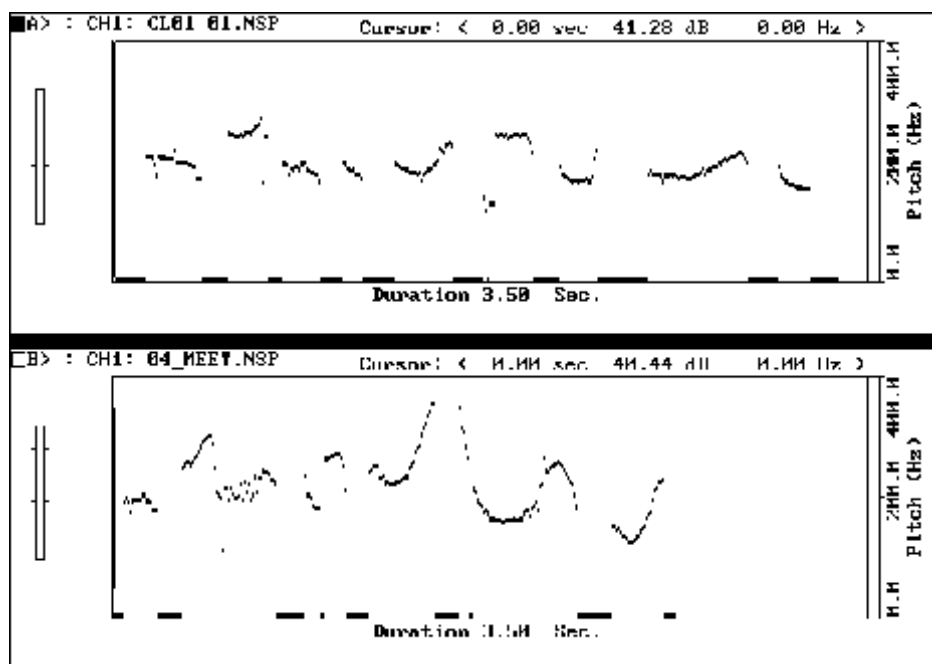
student speaks. After capture, students can play back both the model and their own attempts, seeing the pitch contour in synchrony with the spoken playback, thus reinforcing as many times as required the auditory and visual feedback. This provides instant feedback on unsatisfactory suprasegmental features of student speech.

In addition, on-screen displays or hard copy print-outs can be obtained of various statistical values of the student's and model's speech. These can be used to draw attention to global and local features of the students speech.

The visual feedback generated

In this section, I will exemplify by means of examination of print-outs of a number of the pitch contours generated during the project common features of student pronunciation. It will be seen that on occasions students were unable to approximate to the model's suprasegmental pattern with any accuracy while on others more success was achieved.

Figure 2: Target utterance (TU): // r I'm a**FRAID** // p I have to go to a **MEE**ting // r on **WED**nesday //



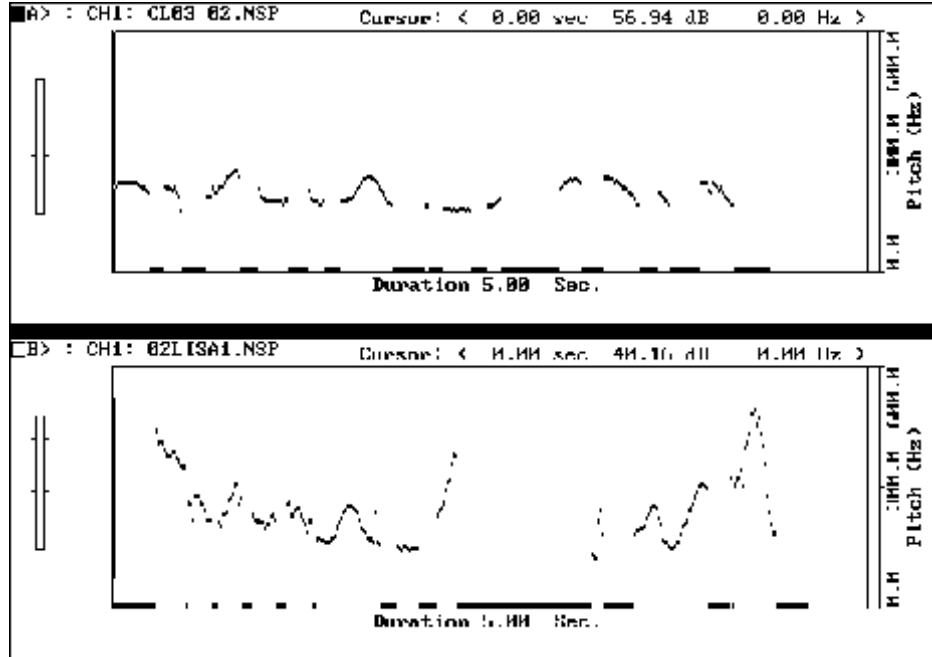
This figure shows the bouncy intonation common to much student speech, in which each high peak reaches much the same level. It will be seen that the uniquely high F0 peak on the word "meeting" as spoken by the model is not replicated successfully by the student, whose pitch rises on each content word. This lack of variation in F0 leads to the perception that no prominence has been given to any particular word and thus to the impression that the speaker has not reacted to or is not aware of the context of interaction.

In addition, the word "Wednesday" spoken by the model with a clear fall-rise tone, is spoken by the model with a fall. This is a very common feature in real student discourse and again gives the impression that the speaker has not reacted appropriately to the situation, in which the idea "Wednesday" is clearly common ground between speaker and hearer and thus should carry a referring tone.

Finally, although not a matter of suprasegmental features, it will be noted from the phonetic transcription that the learner makes a number of errors at the segmental level. Using the editing features of the CSL-Pitch, students can easily highlight such errors.

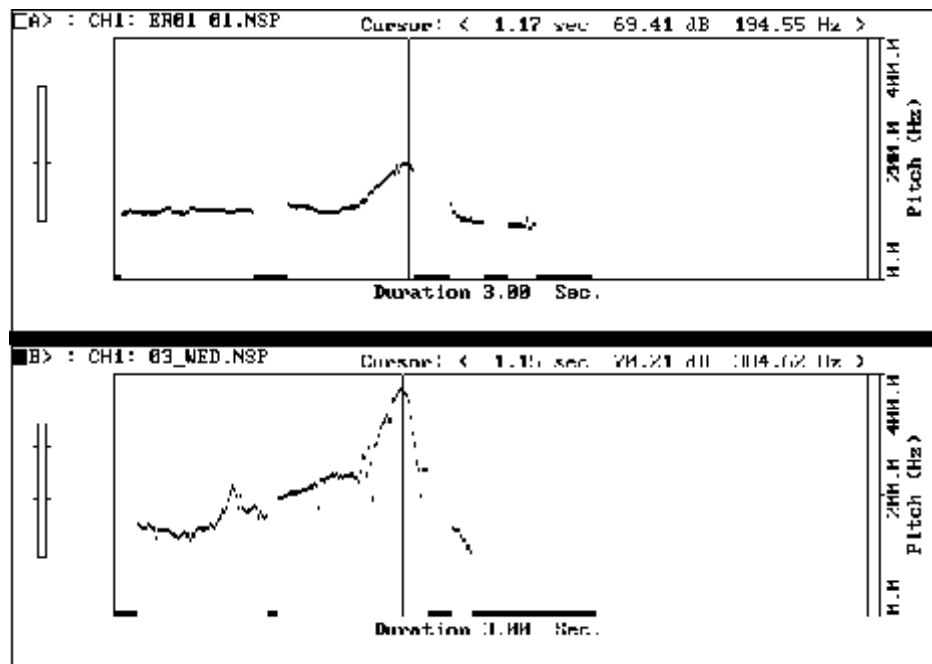
It is this kind of speech, in which segmental and suprasegmental inaccuracy combine, that leads to poor performance in interactive situations.

Figure 3: TU: // p **WELL** // there's no point in **WORRY**ing about it // r what's **DONE** // p is **DONE** //



A related problem is seen in Figure 3, where the student fails to produce enough F0 variation, this time leading to a not incorrect but nevertheless intonationally different pronunciation of the idiomatic phrase "What's done is done". The student's low falling tones give an impression of fatalism, of the utter hopelessness of the situation, whereas the model's higher, livelier-sounding intonation is indicative of optimism, of "putting the past behind one". Again, there are a number of mispronunciations at the segmental level, which could hamper communication.

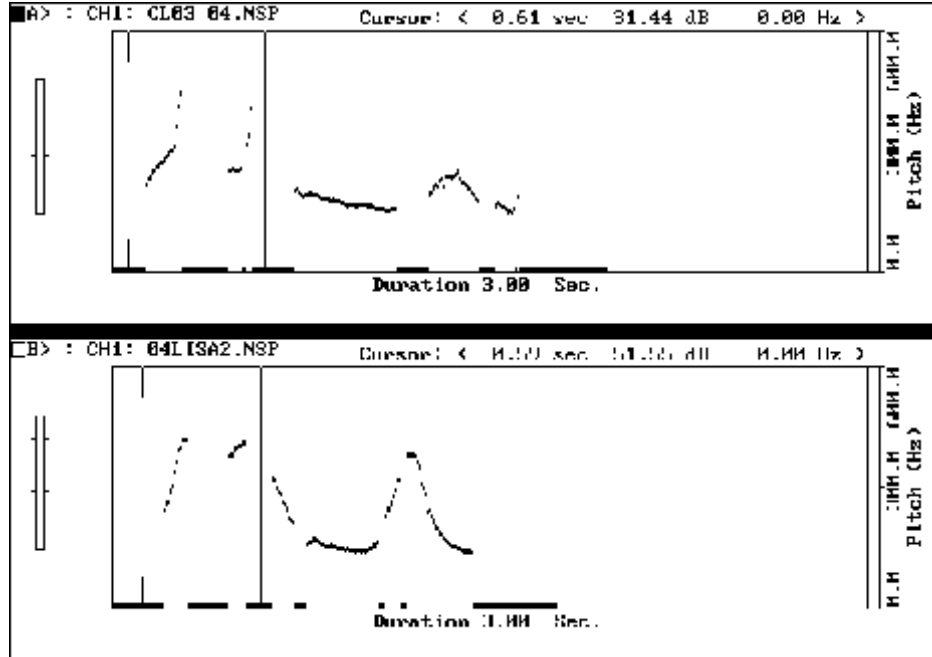
Figure 4: TU: Why don't we go on Wednesday then?



In this example, the student's vocal range, from 81 Hz. to a peak of 194 Hz. at the cursor, is clearly insufficient to bring out the prominence on "Wednesday". In contrast, the model's pitch range is much broader, from a low of 105Hz. to a peak at the cursor of 384 Hz.. Such speech as that of the student sounds bored and certainly fails to bring out the lively enthusiasm intended in the example.

Negative impressions which a learner's speech may give, such as boredom, rudeness, or a failure to react to the situation appropriately are a primary reason for work on suprasegmental features of pronunciation, for they are more subtle and pernicious than shortcomings at the segmental level, interfering as they do on a subconscious level with cultural and social expectations.

Figure 5: TU: // what **DID** you wear //p **ANy**way //



Just such a subtle effect is shown by the contour of the student's utterance in Figure 5. Here, although the F0 reaches a high enough peak, the contours of the pitch on both the words "what" and "did" rises extremely sharply; in the model's version only the word "what" has a comparably sharp rise. This sharp upwards moving contour is characteristic of Cantonese pronunciation of monosyllabic words with a final stop consonant and could, in an utterance such as this, give a listener an unintended impression of brusqueness or rudeness. It is thus on sociolinguistic grounds an area of concern and deserves the students' and teacher's attention. Again, without a facility such as the CSL-Pitch it is extremely hard to isolate such features.

In contrast to this, Figure 6 on the next page illustrates a strikingly successful student modelling of the utterance "Why don't we go on Wednesday then", showing a high pitch peak on the stressed syllable of the word "Wednesday", which, at 373 Hz, is almost the same as that of the model (385 Hz) and follows an almost identical contour. This example shows the positive feedback the CSL-Pitch provides upon satisfactory suprasegmental pronunciation.

Figure 6 (TU): "Why don't we go on Wednesday then?"

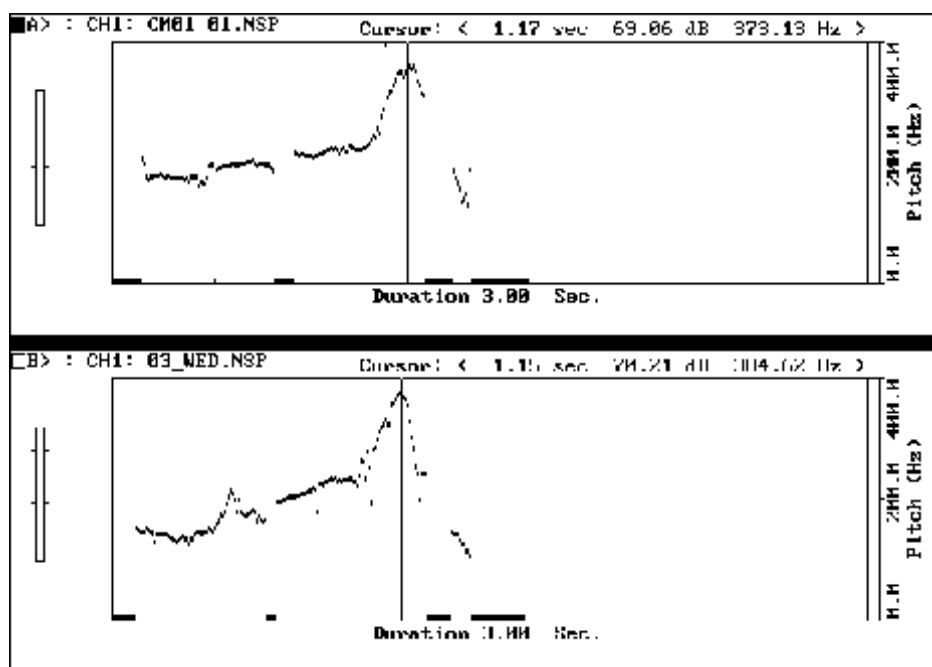


Figure 7: TU: // I MANaged to answer all the QUESTions //

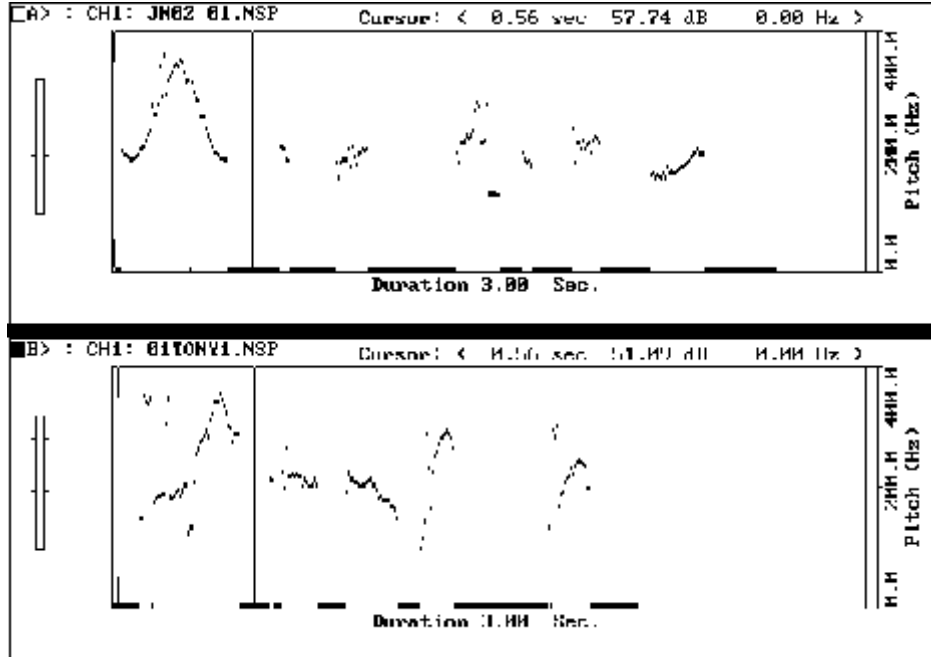
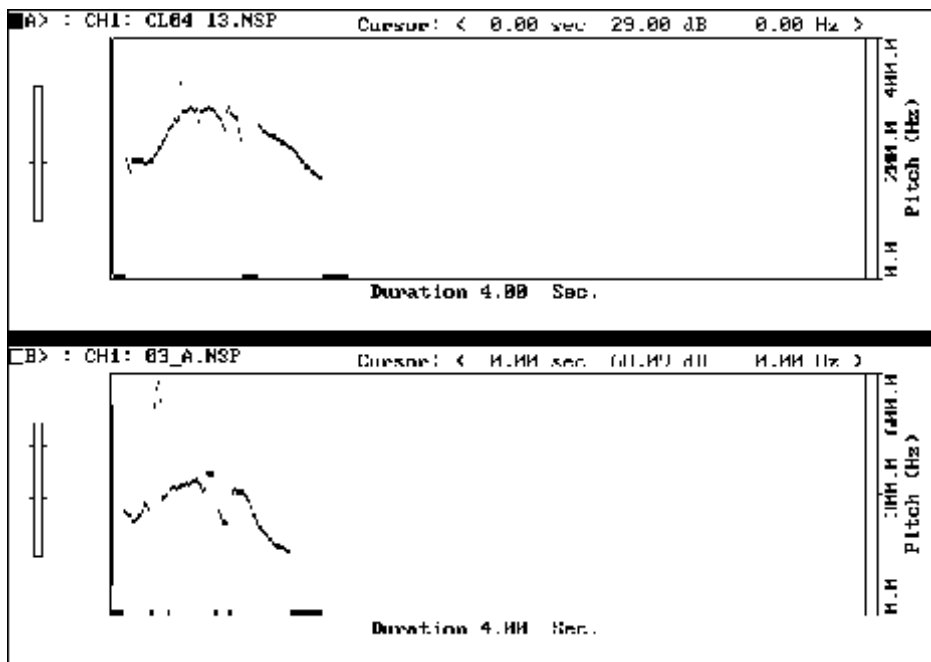


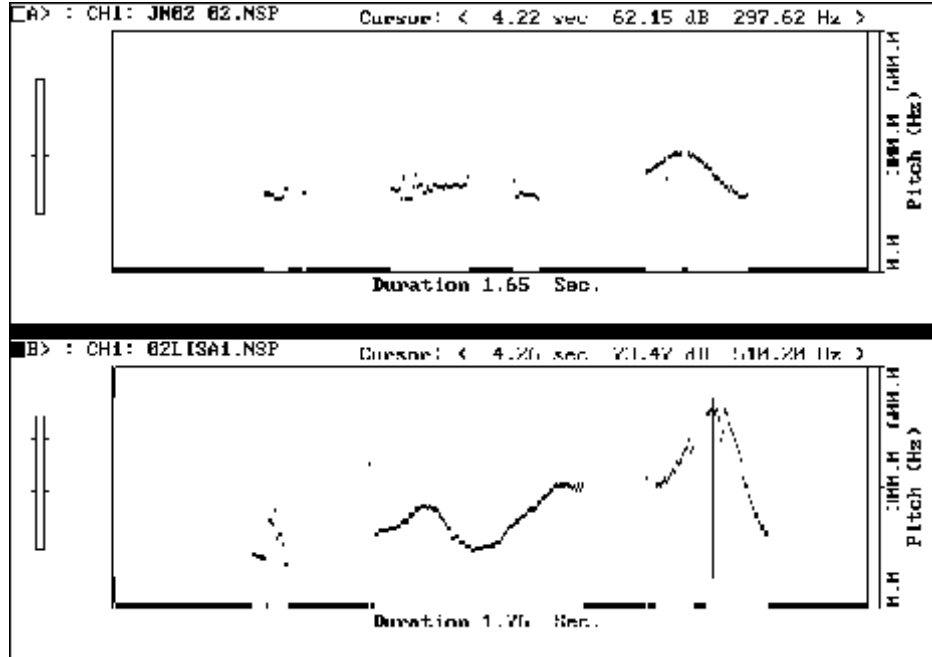
Figure 7 too shows satisfactory prominence on the word "managed", with a high pitch peak and very similar contour pattern, although it will be noted that the pitch contour on the tonic syllable does not match that of the model.

Figure 8: Target Utterance: // i RANG the **BELL** //



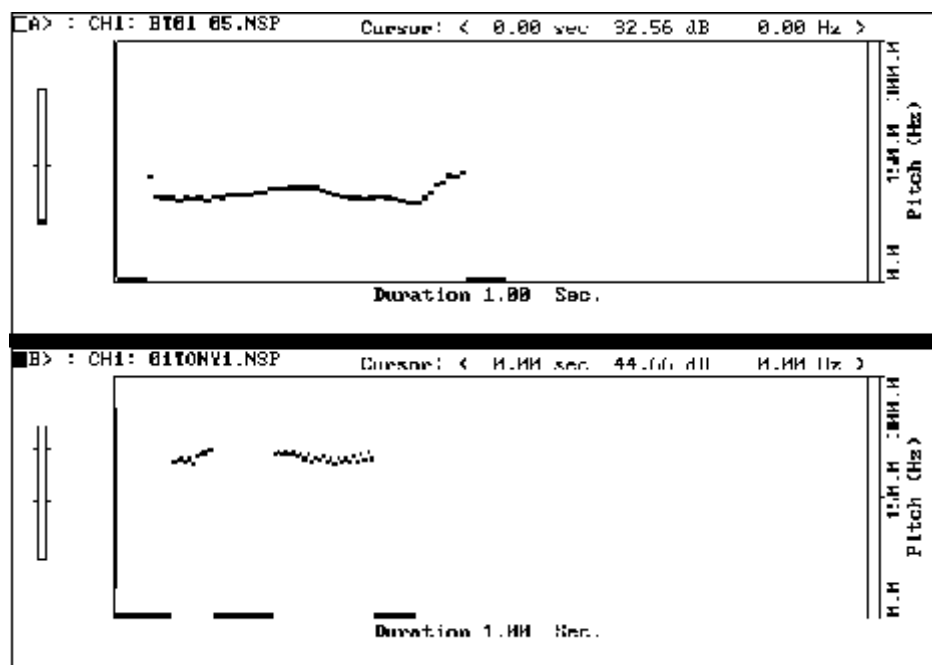
This example shows good control of F0 and duration. The falling tone, as illustrated here, generally causes fewer difficulties to students than do the other tones and so provides a good introduction in the early stages of intonation practice while students are experimenting with the CSL-Pitch and with how to control their own voices.

Figure 9: TU: // r what's **DONE** // p is **DONE** //



Notice the student's maximum pitch peak, at 297 Hz., as opposed to that of the model at more than twice that height. It is most common to find that students' maximum pitch reaches only about half that of the model.

Figure 10: TU: // p I MEAN //



In this example, the student takes 30% longer to say the words "I mean" than does the model and also makes a slightly exaggerated fall-rise tone instead of a falling tone.

Figure 11: TU: // r COULD I have a **WORD** with him //

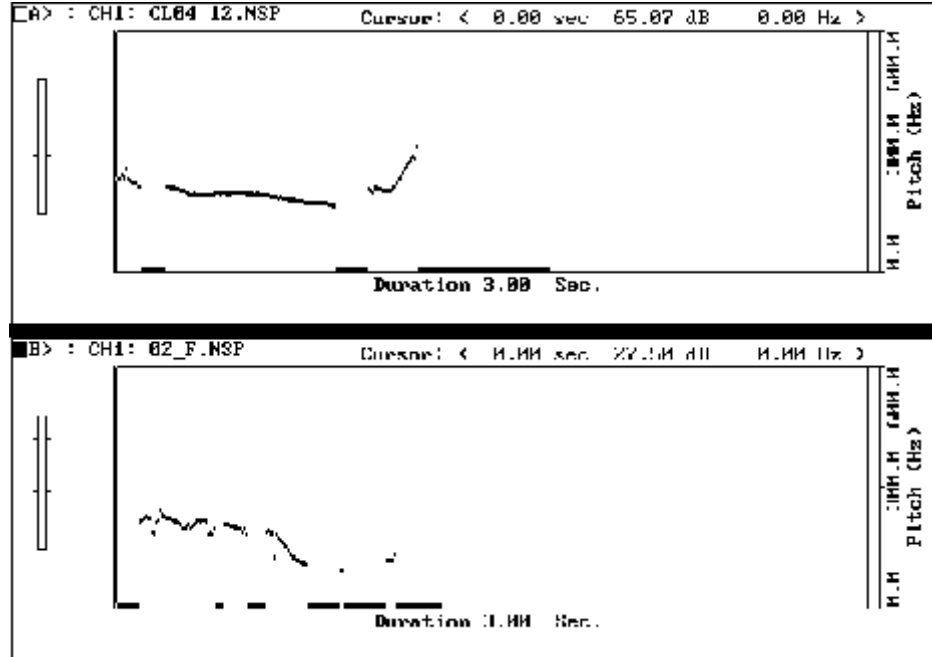


Figure 11, above, shows a more noticeable failure to match the intonation contour, resulting in a very exaggerated fall-rise tone and greatly exaggerated duration on "word".

Figure 12: TU: // r It SHOULdn't be LONG now //

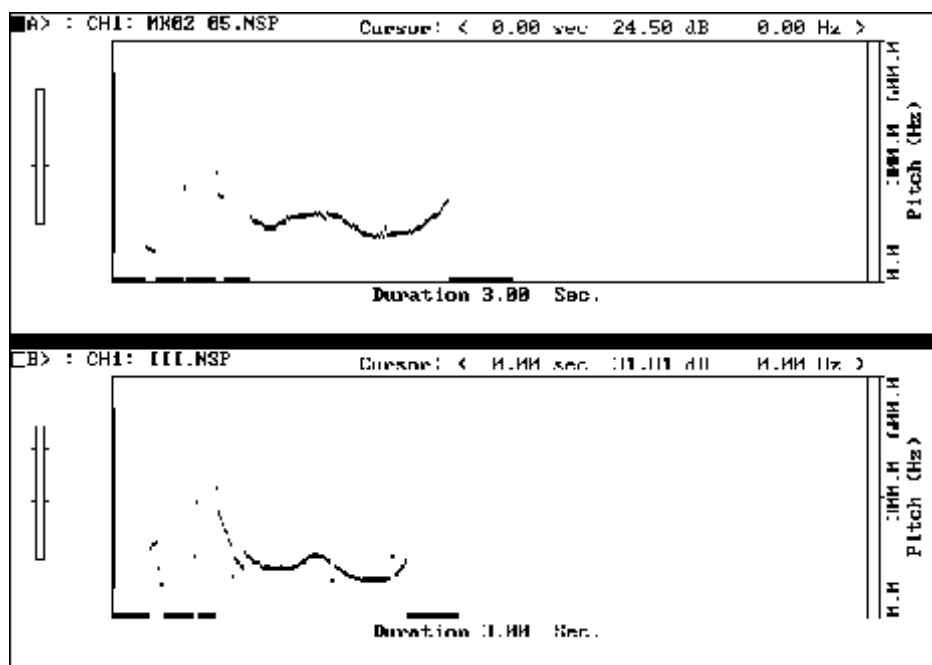


Figure 12 shows an exaggerated fall-rise tone exacerbated by undue duration on the words "long now".

These print-outs are thus examples of the features of student pronunciation which can be examined with much greater certainty than is possible without such equipment.

Pedagogical factors

For the CSL-Pitch to be successfully exploited there a number of factors which must be taken into account which will be discussed here.

1. The CSL-Pitch provides feedback only on short utterances. The maximum possible is 15 seconds but this is much too long to give a clear display. In practice it is more reasonable to think in terms of a maximum of 5 seconds. There is thus a danger of exercises becoming mechanical drills divorced from real discourse. To avoid this happening it is essential that the CSL-Pitch

practice be integrated smoothly into a discourse approach, in which the student is always aware not only of the phonetics of what s/he is saying but also of the appropriate conversational or other interactive setting for it.

2. For the CSL-Pitch to be successfully implemented in a learning programme it must be made available either on a self-access basis or for integration into classroom teaching. If it is to be used in classroom teaching, the problem of sufficient access arises. While it is helpful when using such equipment for students to work together in small groups and thus benefit from each others' help (both technical and phonological), the number who can work effectively on the machine at one time is limited. From my experience, three is the maximum desirable in a group and even then there can arise problems of over-crowding; the student in the centre often has to work the controls for the other two, causing difficulties such as synchronising the recording start time with the speaker's voice. An optimal number would be two students together, but this is not a cost-effective option for classroom teaching.
3. I had expected that a self-access centre would have been established early in the period in which this project took place. As this did not happen, a problem of room space arose, forcing students to use the CSL-Pitch in a noisy room in groups of threes. There are three possible solutions to this problem of access. First would be to install one CSL-Pitch on a stand-alone PC in a self-access area. Supporting this is feedback from the students on the subject, all of whom who used the CSL-Pitch recommended that a CSL-Pitch should be installed in the Self-Access Centre and all but one of whom believed that it was simple enough to be used by students without supervision.
4. The second option would be to introduce the CSL-Pitch into classroom teaching. This would facilitate a more structured approach but would detract from the encouragement of learner autonomy and enhancing self-monitoring skills. In order to achieve a cost-effective and workable student-machine ratio, the CSL-Pitch would have to be linked to a small network with at least eight PCs (for a class of sixteen students) running off a central CSL-Pitch console.
5. The last option is to combine the upper two by installing a network of PCs linked to the CSL-Pitch in an area available for self-access work, so that enough work-stations are available for classroom use while at the same time the area is open for self access work. This last option is undoubtedly the most flexible and has much to commend it on practical grounds.
6. Adequate explanation from the teacher will be needed of the display, what it represents and the physical correlates of the acoustic signal. The games programs, such as "Hummingbird" for pitch control, supplied with the CSL-Pitch provide amusing and effective practice in this. As Molholt (1988: 111) points out, the instructor should also make use of science and engineering students' expertise. In my group there were two final year science students who understood the displays and were able to explain the principles of acoustic analysis to their fellows.
7. The settings of the machine will greatly effect the display. A large number of parameters can be changed to give different results and various different displays are possible, not all of them pedagogically useful. These settings are controlled by a text configuration file. In order to prevent accidental alterations to the settings, the configuration file should be saved as read-only and a back-up copy made and kept away from the self-access area so that the original settings can be easily restored.
8. Of crucial importance is the placing of the microphone, which must be consistent and appropriate. A good position will need to be discovered by trial and error but a general guide is to hold the microphone close the mouth but not so close that loud noises such as sibilants or loud vowels cause overloading.
9. Finally, and of the utmost importance, is that the CSL-Pitch is located in a room free from extraneous environmental noises such as aeroplanes, air-conditioners and banging doors.. Noises such as these are all recorded by the machine and can greatly distort the feedback provided. In fact, this project was severely hampered throughout by such environmental noise.

Conclusions

Current pedagogical theory and the expansion of the higher education institutes is placing increasing emphasis on the development of learners' self-monitoring skills. At the same time, technological advances such as computerised facilities are aiding these developments. Provided that such work is carried out in a way which balances controlled and freer practice, so that drill-work does not dominate excessively, the CSL-Pitch can play a valuable role in enhancing such learner autonomy in an area of English pronunciation which causes particular difficulty to Cantonese learners.

Bibliography

- Bolton, Kingsley and Kwok, Helen (1990) "The Dynamics of the Hong Kong Accent: Social Identity and Sociolinguistic Description" Department of English Studies and Comparative Literature U Hong Kong, Hong Kong *Journal of Asian Pacific Communication*; 1990, 1, 147-172
- Brazil, D. (1985) *The Communicative Value of Intonation in English* Discourse Analysis Monograph No. 8 ELR University of Birmingham
- Brown, Adam (1995) "Minimal pairs: minimal importance?" in *ELT Journal* Vol. 49/2, April 1995, 169 - 175
- Coulthard, M. (1992) "The significance of intonation in discourse" in Coulthard, M. (ed.) *Advances in Spoken Discourse Analysis* Routledge London
- Cruttenden, A (1986) *Intonation* CUP Cambridge
- Jones, R.H. and Evans, S. "Teaching pronunciation through voice quality" in *ELT Journal* 49: 3 244- 251
- Fry, D.B. (1979) *The Physics of Speech* Cambridge University Press
- Lambacher, S.G. (1996) *Spectrograph Analysis as a Tool in Developing L2 Pronunciation Skills* Feb. 28, 1996
<http://www.u-aizu.ac.jp/~steeve/speakout2.html> (July 1, 1996)
- Leather, Jonathan (1983) "Second-language pronunciation learning and teaching" in *Language Teaching* July 1983, Vol. 16, no. 3, 198 -219
- Molholt, G. (1988) "Computer-assisted instruction in pronunciation for Chinese speakers of American English" in *TESOL Quarterly* Vol. 22, No. 1, March 1988, 91 - 111
- Morley, Joan (1991) "The pronunciation component in teaching English to speakers of other languages" in *TESOL Quarterly* Vol 25 (3), 481-520

Throughout this paper the example utterances are from Brazil 1994 and Bradford 1988.

The Internet TESL Journal, Vol. II, No. 8, August 1996

<http://iteslj.org/>

<http://iteslj.org/Articles/Stibbard-Intonation/>