Articulatory setting and voice production: Issues in accent modification

Introduction

In an interesting article discussing the difficulties that students with a Vietnamese language background have with English pronunciation, Macneil (1987) suggested that a significant factor is the transfer of a staccato speech pattern related to a breathing style which is different in Vietnamese and English. She argued that, as English is a multisyllabic language, it makes extensive use of linking between words and requires the maintenance of an even air pressure through long utterances, while Vietnamese, which is an essentially monosyllabic language with few final consonants, utilises short pulses of air-pressure that cut off before the end of the syllable (Macneil 1987: 64).

Although Macneil did not provide any objective evidence for her hypothesis, and so it must remain speculative, her work, nevertheless, serves to remind us of the importance of speech mechanisms other than the lips and tongue, and suggests that, as teachers, we may need to deal with some fundamental aspects of the way in which the voice is used in speech in order to assist learners whose accent poses a real barrier to intelligibility. This article briefly describes a single case study based on this principle and discusses some of the implications for classroom practice.

The vocal tract and vocal resonance

The vocal tract comprises the lungs, the trachea, the larynx or ‘voice box’, the pharynx and nasopharynx, the mouth, tongue, lips and jaw. Figure 1 illustrates the major areas of the vocal tract above the lungs.

Breathing is the mechanism that drives the speech production system. The breath from the lungs passing through the larynx creates vibrations of the vocal folds. The sound produced by the vocal folds themselves, however, is tiny. What we hear as a speaker’s voice is caused by the resonating of this tiny sound in the pharynx, the oral cavity, and the nasal cavity. It is this resonance that gives the voice its carrying power. The resonance of a particular voice depends partly on individual anatomical characteristics such as the size of the resonating cavities, but also on the way in which the resonating cavities are used. Changes in the size and shape of these cavities
occur as the lips, tongue, jaw, soft palate and pharyngeal muscles move during speech. The movements of the soft palate, for example, increase or decrease the nasality of the voice. The forward or backward positioning of the tongue causes the sound to be reflected either from the hard surfaces at the front of the mouth — **anterior resonance** — or from the soft tissues at the back of the mouth and in the throat — **posterior resonance**. A posterior resonance tends to produce a more ‘muffled’ sound because the resonating surfaces are softer than the hard palate and the teeth (Moncur and Brackett 1974; Fry 1979; Andrews 1995).

![Figure 1: The vocal tract (adapted from Weiss 1978: 27)](image)

As a speech pathologist and ESL teacher, I found that a very striking characteristic of students from a Vietnamese or a Cantonese background was the quality of resonance in their speech. When the students spoke their own languages, the voice seemed to me to resonate in the back of the mouth and the nasal cavity. This pattern appeared in their English as well. Voices of native speakers of English, however, appeared to resonate more in the front of the mouth. An informal survey of a group of other speech pathologists confirmed that they had the same impression. When a videotape
of a range of speakers of different language backgrounds was played to this group, posterior resonance was commented on for Vietnamese, Cantonese and Korean speakers, but not for speakers of German or Turkish.

If it is the case that there is a difference in the resonance patterns of, for example, Cantonese and English speakers, it would seem logical to infer that this difference is in some way related to the articulatory movements used in producing the sound combinations of these different languages. The relationship between the sounds of a language and the quality of a speaker’s voice was referred to by Honikman (1964) as ‘articulatory setting’.

**Articulatory setting**

Articulatory setting does not refer to the specific movements needed to produce individual phonemes, but to habitual movement patterns. The articulatory muscles are held in a certain pattern of readiness to facilitate the most efficient movement from one sound to another in the flow of speech. Honikman argued that the most frequently-occurring sounds and sound combinations of a language largely determined the articulatory setting. For example, English, which has a high frequency of alveolar sounds produced with the tongue tip or blade, tends to favour a setting that makes rapid anterior tongue movements easier. She argued that it was not possible for a speaker to master the pronunciation of a second language (L2) completely if they maintained the articulatory setting of a language with a dissimilar phonological system. The work of Honikman has been extended into a wide-ranging study of the changes in voice quality that result from different articulatory settings, notably in the work of Abercrombie (1967), Laver (1980), Esling and Wong (1983) and Esling (1987, 1994). Some of these authors use the term ‘voice quality setting’ in preference to ‘articulatory setting’.

**Changing focus of resonance**

In a single-case study, I used a speech pathology technique known as ‘Changing Focus of Resonance’ (CFR) in an attempt to improve the intelligibility in English of a speaker of Cantonese language background (Kerr 1998, 1999). CFR involves using the kinaesthetic vibratory feedback provided by the bones of the head and face to ‘place’ the voice for maximum resonance for the sounds of the English language. The subject is taught to attend consciously to the kinaesthetic feedback in the cheekbones, nose, lips, teeth and their supporting bones, at first in slow, prolonged speech, and gradually in speech of normal conversational rate. Important aspects of the technique are a relaxed tongue, a widened oral cavity, and a ‘gentle onset’ which reduces tension in the vocal folds and pharynx so that the sounds are initiated smoothly rather than sharply (Moncur and Brackett 1974; Fawcus 1986). Enlarging the oral cavity and relaxing the vocal folds and pharyngeal muscles foster an anterior focus of resonance, rather than
the posterior resonance that has been hypothesised to be more common among speakers of some Asian languages.

The CFR program followed in this study consisted of twelve individual 75-minute sessions. It began with a very simple explanation of the notion of articulatory set, and with the highlighting of two principles: that English uses ‘the front of the mouth’, and that it is essential to produce the vowels of English clearly. Vowels and diphthongs were targeted in the CFR program because they are the most prominent part of the English syllable (Fry 1979: 126; Fromkin, Rodman, Collins and Blair 1990: 43; Pennington 1996: 89). Vowels occupy the major part of the time in any stretch of speech, and are the loudest sounds. Because the acoustic cues provided by vowels are more prominent and long lasting, vowels carry the bulk of the prosodic information; pitch and stress information is carried on the vowels (Ladefoged and Maddieson 1990; Pennington 1996: 89–90). Second-language speakers who produce vowels with resonance patterns closer to those of a native speaker may, therefore, also use patterns of prosody closer to those of a native speaker.

In the first session, vowels and diphthongs were imitated in isolation. The subject was instructed to produce the vowels with the tip of the tongue behind and touching the bottom teeth to discourage bunching-up of the back of the tongue, which creates a constriction at the back of the mouth (Moncur and Brackett 1974: 54). Differences in vowel sounds were explained with reference to lip shape and range of jaw movement, not tongue positioning, so that the subject was concentrating on enlarging the oral cavity by keeping the tongue low and using a wider range of jaw movement, and on the most anterior part of the articulatory apparatus, the lips.

The second session introduced the concept of resonance by using the initial stages of the approach outlined in Fawcus (1986:152). The subject, CC, began by cupping his hands around his mouth and nose, and intoning \textit{bm-bm-bm-bm} on a continuous sung tone. The combination of the labial and the nasal gives strong vibratory feedback from the lips and the bones of the front part of the face, the hard palate, and the nasal cavity. Cupping the hands around the mouth and nose intensifies this kinaesthetic and vibratory feedback so that the subject is more aware of it. CC was asked to describe where he could feel the ‘buzz’ or vibration, and when it was clear that he could detect this, he was asked to repeat the exercise with his hands held still in a cupped position, but approximately 10–15 cm in front of his face. The cupped hands acted as a reminder to try to place the voice as far forward as possible, but in this exercise CC was attempting to attend to the kinaesthetic feedback within the cavities without the tactile feedback from the hands. When CC reported that he could feel the kinaesthetic feedback in this way, he then moved on to intoning vowels and diphthongs and reporting on the different feedback from different vowels — for example, the increased ‘buzz’ felt in the teeth for /i/ as opposed to /u/, and the
increased feedback from the lips for the lip-rounded vowels. Most of the emphasis was placed on the more anterior vowels and diphthongs so that the concept of keeping the voice forward was maintained.

In this and the following session, vowels and diphthongs were practised in a slightly prolonged form in nonsense syllables, both open syllables — consonant + vowel — and closed syllables — vowel + consonant (Moncur and Brackett 1974: 167). Nonsense syllables were used initially to avoid any influence from the subject’s learned patterns for pronouncing familiar words, and gradually real words and short phrases were introduced. The emphasis in these sessions was on the concept of frontal resonance and the kinaesthetic feedback experienced by the subject.

Sessions 4 and 5 addressed the issue of the different English spelling forms a vowel or diphthong may take. A potential source of pronunciation error for many L2 speakers is the complex nature of sound/spelling correspondences in English, and this may be compounded by the fact that L2 instruction in the learners’ own country is often focused on written texts rather than on listening materials using native speakers (Acton 1984; Browne and Huckin 1987). A list of the most common spelling forms for each vowel or diphthong was given (based on Fromkin et al, 1990: 53–4), and the words given as examples were practised. The text of a popular song was then given to the subject. I read the first line at a slightly slower rate than normal conversational rate, but maintained the normal vowel values for connected speech. For example, the line ‘the long and winding road/ that leads to your door’ was read as /ðəl əŋ ən waindiŋ roud / ðət lɪːdz ə ʃɔːd/ with the weak vowels rendered as /ə/. A song was chosen for this task to encourage production practice which was smooth and rhythmical with liaison between words, but which was at a slightly slower rate than that of conversational speech. The subject was asked to identify the vowels or diphthongs in each word, write in the phonetic symbols, and then repeat the line. If I felt that the anterior resonance was not being maintained, I instructed the subject to ‘bring the voice forward’ and used the cue of putting cupped hands slightly in front of the face. Auditory feedback was introduced at this stage: the session was taped, and I stopped the tape and played back any section in which I felt the resonance was not being maintained. CC then repeated that section and the tape was played back to demonstrate to him the change in loudness and clarity of the voice when he used anterior resonance.

In the following seven sessions we worked on incorporating the resonance technique into CC’s conversational speech, using tasks of increasing linguistic complexity. Work-related terminology in single words and short phrases was used initially. He identified the vowels and diphthongs in what I said to him, wrote down the phonetic symbols and practised the words and phrases, slowing the speech rate slightly by drawing out the vowels. I cued him to maintain liaison between the words in the phrase with the phrase ‘keep the voice going’. Conversational tasks using the terminology
that had been practised were then introduced. CC was interviewed about his work and his personal history and I constantly gave him feedback in the form of the correct version of words in which he had made a vowel or diphthong error, through cueing him to produce a more anterior resonance when needed, and to ‘keep the voice going’. Reading exercises using work-related terminology were also used at this stage to give him practice in concentrating on anterior resonance without having to think about vocabulary and grammar at the same time.

The linguistic demand on CC was gradually increased by requiring him to produce a short, unprepared answer to a question about some aspect of everyday life. For example, I asked him, ‘What would you do if you lost your credit card?’ These questions were based on exercises contained in Martinoff and Stokke (1981) and Lazzari and Peters (1991). I cued him where necessary using the cupped hands gesture, and I played the tape back so that he could assess for himself whether he was maintaining anterior resonance and ‘keeping the voice going’. The last two sessions involved longer conversational exchanges on topics such as smoking, driving, diet and exercise, and superstitions. These topics were taken from exercises in West (1996).

Implications of the study

An empirical investigation of the efficacy of this program (see Kerr 1998, for full details) raised some interesting issues. Although the study design did not allow the conclusion that the CFR program had improved CC’s intelligibility, there was a strong indication that native-speaker judges were better able to understand longer sections of utterances. The Appendix illustrates this point.

Native-speaker judges appeared better able to understand longer stretches of speech, even though there were no significant changes in segmental accuracy as investigated measured using the Weiss Articulation Test (1978). Although the study was not able to conclude why this was so, one possible explanation is that there was a change in CC’s prosody, specifically in the salience of stressed syllables or the ability to sustain voice over longer phrases, both of which could be predicted as an outcome of the CFR technique.

Several observations made during the course of the study may be of interest to ESL teachers. It seemed to be easier for CC to attend to relatively gross articulatory movements such as general tongue relaxation, jaw opening, and lip rounding or spreading, than to more complex articulatory patterns in which tongue movements and postures are finely specified. In addition, as Acton (1984) argued, kinaesthetic feedback may be more accessible to learners than is auditory feedback, particularly when dealing with a learner with entrenched error patterns that may be related to a faulty auditory model of the sound.

The fact that the program takes a cognitive approach may also be important. The rationale for the technique was explained to CC, and he
was provided with a terminology which he and the researcher could use to reflect on his performance. Thus CC was actively involved in modifying and monitoring his performance, and was given the tools with which he could do this independently of me. This appeared to give him confidence in his ability to speak more clearly when he consciously attended to the technique. This methodology may be particularly effective for adult learners (Browne and Huckin 1987; Anderson-Hsieh 1990; Morley 1992).

Field-specific materials (Anderson-Hsieh 1990) also seem to have been an important factor in helping CC transfer the principles of CFR quickly to more personally meaningful conversational exchanges. Further, it was clear during the sessions that many of CC’s mispronunciations of words related to his work or studies were based on the spelling of these words. For example, CC pronounced ‘research’ as /rɪsɪʃ/, but when he was required to listen to my model, identify the vowel, and write in its phonetic symbol, he was able to modify his production of the word and to use the new pronunciation.

**Pedagogical implications**

Although the findings from this case study were suggestive rather than definitive, several pedagogical implications emerged. Firstly, the study suggests that global improvements in intelligibility may be made by working on vowels alone. If this work is allied with focus on a field-specific vocabulary, it may be possible to produce a considerable improvement in functional intelligibility, that is, in intelligibility in certain prescribed situations such as the workplace, even when the learner continues to have difficulty with unfamiliar vocabulary. Secondly, there was an interaction between CC’s pronunciation difficulties and his command of syntax and vocabulary. In the course of the study, it became clear that CC was better able to concentrate on the CFR technique in situations where he did not have to formulate complex responses. This is not surprising, given the cognitive load involved in attending simultaneously to accessing vocabulary, organising syntax, and motor planning for the movements of the articulators. The implications of this are that we need to make pronunciation an integral part of the teaching of new grammatical structures. In other words, as a new grammatical structure is learned, it may be important to give the learner extensive practice in identifying the vowels in the stressed syllables and producing the structure with the correct vowel qualities. On the CFR program, structures were practised first in slower, more prolonged utterances. The speed of utterances was gradually increased to normal speaking speed. This progression may also be important in helping the speaker internalise the motor patterns. It is arguable that failure to pay sufficient attention to the motor patterns of an item may result in the learner’s ‘learning’ the grammatical structure theoretically rather than ‘acquiring’ it in spoken performance.
CC very often did not know how to pronounce words that were important for his work, even though he had seen them written, and presumably heard them spoken many times in the course of his studies. When he worked with my model to identify the main vowels, he demonstrated that he was able to pronounce many of these words acceptably, and to maintain this pronunciation. This implies that saying the word with the correct vowel qualities may be an essential component in the acquisition of new vocabulary items. All too often learners are focused on looking up new words and writing down their meanings rather than listening to how they are actually said and developing stable and recognisable productions. An essential component of this work, for both teacher and student, is a confident grasp of the phonetic symbols used for vowels and diphthongs. Given the difficulties of English spelling patterns, we need to give learners some reliable tool with which they can impose some order on the sounds they hear and attempt to reproduce.

**Conclusion**

This study contributes to our understanding of the way in which the vocal tract works as a whole to produce spoken utterances, and suggests that attention to voice quality has the potential to improve overall use of prosody. Jenkins (1997) proposes that today, more than ever, we need to be thinking of English as an International Language (EIL) rather than as a second or foreign language. Hence, our focus needs to be on intelligibility rather than on particular norms of native-speaker accent. The three areas Jenkins proposes as the most important for intelligibility are ‘core sounds’ that occur frequently and across all varieties of English, the main stress in a word group, and the effective use of articulatory setting (Jenkins 1997: 122–3). This case study suggests that, even in the presence of numerous segmental errors, changes to articulatory setting may produce a noticeable improvement in intelligibility.

**Appendix**

The texts of sentences randomly extracted from conversation are shown below. The italicised portions indicate the sections of each utterance that were correctly transcribed by at least three of the four judges.

**Baseline 1a**

1. My home/ er, usually I/ talk the Chinese language
   
   *Usually (3/4 judges)*

2. Mm, each placement is the three-week.
   
   *Three weeks (all judges)*
3. I have learning/ I have been learning the course of the Disability Study
   I have been learning..........disability (all judges)

4. You are/ you can cope with something
   You are, you can (3/4)
   Something (all)

5. Such as the/ mm, the hostel/ and the/ adults’ day/ service and support
   Hostel (3/4)

6. Because they got a, a/ a lot of experience about the life
   .........................

7. Before I/ work as/ as a doctor in Shanghai
   A doctor ... Shanghai (all)

8. I think this policy not good
   I think (all)
   This policy (3/4)

9. Also I got the/ one dog
   .........................

10. Children, they de, develop/ not good for they
    .........................

11. Yeah, but the letter is the same, the Chinese letter
    Chinese (all)

**Baseline 1b**

1. Usually the half an hour/ for/ each treatment
   .........................

2. Because at the moment I think more difficult
   Because at the moment (3/4)

3. You, you/ didn’t got the/ too much the/ patient
   .........................

4. Oh, er, go to park/ yeah, go to park/ with my dog
   .........................

5. And the only you/ talk about how to/ about the acupuncture point
   About acupuncture (all)
6. Only mm, one one and two years old, very children
   One and two children (all)
7. You can give some advice
   You can (all)
8. You need more time, maybe forty five minutes maybe the one hour
   Forty five minutes one hour (all)
9. And the older one, maybe the eighty, mm, the ninety
   .....................
10. Too much — I don't know the where they come from
   Too much, I don't know where they come from (3/4)

Post-treatment
1. But this, I think not only the Hong Kong, is the whole world
   I think ... not Hong Kong ... the whole world (all)
2. Yeah, I got a car.
   Yeah, I got a car (all)
3. Yeah, go to work and go to shopping, and go to visit the friend.
   Go to work, go shopping ... and to visit (all)
4. And sometimes we go together, go to park
   Sometimes we go together, go to park (all)
5. Mm, I think the a lot of Hong Kong people don't like the communists
   I think a lot of Hong Kong people don't like ... (all)
6. My house, go to station is the ten minutes about
   To station (3/4)
7. 1985 I find the people more polite
   1985 I find people more polite (all)
8. He, he he know he is wrong
   He know he is wrong (all)
9. Yeah, they use for the, transport the food and the, the business
   Transport (all)
10. Yeah, the, the people live in the China is very very poor
   The people in China is very poor (all)
References


Ladefoged, P and I Maddieson 1990. ‘Vowels of the world’s languages’. *Journal of Phonetics*, 18, 1: 93–122


