prpn SIG

SPEAK OUT!

THE NEWSLETTER OF THE IATEFL PRONUNCIATION SPECIAL INTEREST GROUP

September 2008 Issue 39

Price £4.00
Free for PronSIG members
ISSN 1026:4345
www.iatefl.org

What if children don't learn to pronounce by imitation? How should we teach older learners?

Piers Messum

We believe that children learn to pronounce by listening to adults and then basing their production on what they have heard. So when we teach older learners, most of the exercises we propose ask them to 'listen and repeat' in one form or another. We think that this approach should be effective because it is consistent with what seems to be the natural way of learning. But taken as a whole, the results we get are not as good as we would hope. Perhaps we are wrong in some ways about what is natural …

I recently completed a PhD thesis that examines the role of imitation in learning to pronounce (Messum, 2007a). It challenges the orthodox assumption that children do this 'by imitation' and proposes alternative mechanisms to account for the replication of some aspects of pronunciation. This is a report of my presentation at the 2008 IATEFL conference, where I described two of these mechanisms, and some of the practical implications for pronunciation teaching if they are indeed those involved.

I do not have the space to present all the evidence in favour of my proposals or to explain why the orthodox assumption is probably wrong. For a more careful and complete argument please see the short articles I have written (Messum, 2008; 2007b; 2007c) or the thesis itself. All of these are available on my website, as are the slides from my presentation at Exeter.

Part 1: learning the timing patterns of English

In the first part of my thesis I considered the modifications in the timing of spoken English that are typically discussed in phonetics textbooks, including the 'rhythm' of the language, the differing lengths of tense and lax vowels in certain contexts, and voice onset time (aspiration). These phenomena are all believed to be time-related not only in perception but also in production, meaning that the speaker is planning his articulation to achieve timing goals rather than to achieve some other objective which might lead to changes in timing but only as an unintended by-product.

The other side of the coin from this is that these phenomena are supposed to be learnt by imitation, i.e. by children noticing the timing changes in the speech of others, working out some set of underlying, time-based 'rules' that capture each one, and then using these rules to guide their own production. The way that we conventionally teach these phenomena to second language learners assumes that they will be able to undertake a similar modelling process.

Let us consider the 'rhythm' of English, which is a challenge for many learners. Others (e.g. Dauer, 1983) have described how a variety of non-temporal aspects of the language contribute to the effect of so-called stress-timing, and they therefore question its reality. But to the best of my knowledge no-one sympathetic to this theoretical viewpoint has explained why foot level shortening (FLS) occurs. This is a key issue, since FLS probably contributes more than any other factor to the impression of stress-timing.

FLS is the phenomenon of a speaker progressively shortening the syllables in a foot as more syllables are added to it. (Compare | *one* | *two* | *three* | with | *one and* | *two and* | *three* | and | *one and then* | *two and then* | *three* |.) One of its effects is to make the intervals between stressed syllables more isochronous than they would otherwise be, so in the absence of any other motive FLS has seemed to be evidence in favour of some kind of 'rhythmic' consideration being part of the speech planning process for English speakers.

To explain why FLS occurs, it helps to start with consideration of pre-fortis clipping (PFC), the shortening of the vowels in *cat* and *niece*, for example, compared to their lengths in *cad* and *knees*. PFC is apparently unrelated to either FLS or the rhythm of English, and is almost a universal feature of speech in languages where it can occur.

In trying to explain PFC, what has been overlooked up to now is the fact that a young child's respiratory physiology is very different from that of an adult. At the end of an inhalation an adult generates roughly the subglottal pressure he needs for speech solely as a result of the new configuration of his chest wall. He has stretched his body tissue and its elastic recoil compresses the air in his lungs. But a child's very compliant ('floppy') chest wall means that little subglottal pressure is generated as a result of inhalation. The contrast between the adult's pulmonary-chest wall unit and the child's can be imagined as that between an inflated balloon and an inflated paper bag.

Air doesn't flow out of a paper bag of its own accord. To create the subglottal pressures that drive the airflow for speech a child must use volitional expiratory muscular activity throughout an utterance. (And the pressures he requires are, in fact, higher than an adult's and have to be produced with a weaker musculature.)

For this and for some other reasons, a child's expiratory speech breathing will be pulsatile. Further, because this is a motor skill that must be learned while speech is being developed the degree of respiratory system effort will be largely independent of the activity of the upper articulators whose acoustic consequences it supports. Another way of describing this is to say that each pulse of respiratory system activity will define a fixed 'frame' into which the 'contents' of articulation downstream will be fitted. (Such frame/content relationships are common in the development of motor skills.) The pulse of respiratory system activity that is associated with a syllable will vary, for example, for overall loudness, but not for the different type or number of 'segments' that might be articulated with its support.

PFC is a natural consequence of this model of speech breathing in a child. A fortis consonant always requires more aerodynamic resource than a lenis one, so in a system of syllable production where respiratory system activity is invariant the extra resource needed for, say, a final /s/ as compared to a final /z/ must be found by allocating less resource to previous segments, particularly to the preceding vowel. The child has no timing target for the vowel. When it is 'clipped' it is because this is how he can conveniently satisfy an aerodynamic constraint.

This explains all of the curious characteristics of PFC, including its dependence on stress (De Jong, 1991), its appearance when words are spoken clearly but its absence (typically) in conversational speech (Crystal & House, 1988), and its early appearance in children, from 22 months (Naeser, 1970). Further, it would explain how it is that Swedish children show PFC at 24 months but then drop it by 30 months (Buder & Stoel-Gammon, 2002). Dropping or overriding it is not surprising, since it does not appear in the adult language, where vowel length is phonemic and shortening for other reasons would be problematic. But where did the 'clipping' come from in the first place? The Swedish children did not hear PFC in the speech of others, so they could not have replicated it by imitation.

Notice that the mechanism I have described for PFC also explains another 'timing' phenomenon: the compression of the main vowel in a word as segments are added to it, creating a cluster of consonants (as in *ram, ramp, ramped)*. Here again, as more elements are added for production supported by an invariant pulse of respiratory system effort, so the resource allocated to support each must be diminished and hence the time of individual execution reduced.

It is now a short step to explain FLS without needing to invoke any notion of rhythmicity. The process is again one of distributing limited aerodynamic resource over elements that all require a share of it, and FLS occurs in English because of two of the distinctive features of the language.

Firstly, English employs 'stress-accent', so the mechanism for routinely stressing a syllable (making it more prominent than those around it) leads to an increase in loudness as well as in length. In most languages length and pitch change for sentence stress, but not loudness. If a transient increase in loudness is a goal then adults can produce this via a laryngeal adjustment alone, but children have to increase their respiratory drive, applying more effort in the creation of a pulse[1].

Secondly, the aerodynamics of reduced vowels in English mean that from the point of view of the respiratory system there is no great difference between weak syllables and consonant clusters: both are periods where the vocal tract offers high resistance to airflow. So where from an auditory perspective we say that a foot contains a stressed syllable followed by zero or more reduced syllables, the respiratory system sees everything after the stressed vowel as a single high-resistance unit; in a sense, as a complex consonant cluster. *Fricatives and resonants* would appear as || CCVC·C·CC·C | CVC·C·CCC || where dots represent momentary transitions between one consonant and the next.

We can now see how the mechanism that explained PFC will also come to operate at the level of the foot in English, rather than just at the level of the syllable as in most other languages. English demands a pulsatile style of speech breathing in a young speaker not only for syllables in his earliest speech, but later too in order to create stress prominence. The child then has to distribute an invariant amount of aerodynamic resource available for the foot over the domain of the stressed syllable plus following unstressed ones. The addition of extra weak syllables demands part of this resource that can only be supplied by it being withdrawn from existing syllables, which therefore have to be articulated more quickly.

The result is FLS, whose name, like PFC, correctly indicates a perceptible change in timing ('shortening' or 'clipping') but a change which is not motivated by timing concerns *per se*. The real issue for the young speaker is the need to allocate each quantum of aerodynamic resource created by his pulsatile speech, breathing across all the elements that it needs to support.

[1] Singers will recognise a parallel here with "singing on your cords" and similar expressions that describe using the larynx to achieve vocal effects that are better created with increased activity on the part of the abdominal muscles of the respiratory system (the singing then being described as "well supported").

Part 2: learning the sound qualities of English

In part 2 of my thesis, I looked at how children learn to pronounce and use speech sounds as part of learning to say words. We don't learn the pronunciation of words by copying 'whole-word' shapes, although young children do probably learn to say their very first words this way. Instead, we break words that we have heard down into their constituent speech sounds and then reassemble them using the speech sound equivalents that we have in our production repertoire. So to learn the pronunciation of a new word like *horripilation*, we might parse what we hear into 4 or 5 roughly syllable-size chunks, and then say what we know to be the equivalent of each of these in our own voice. We are learning the pronunciation of this word by imitation (copying the serial order of its elements), but notice that the process relies upon the prior establishment of correspondences between (1) speech sounds produced by others and heard by us, and (2) speech sounds produced by us.

Creating those correspondences is what I think we should call 'learning to pronounce' (as opposed to 'learning to pronounce words'). It has always been assumed that learning to pronounce also happens through an imitative mechanism: children comparing a speech sound they produce to what they hear, and adjusting the former if they judge it to be dissimilar. In the thesis I argue that this is not possible. That's initially a very surprising and counterintuitive assertion, but there is more than one good reason for it to be true, particularly in the case of the vocalic part of speech sounds.

But if it's not by copying speech sound qualities by ear, how can young children learn to pronounce speech sounds? The key interaction, I believe, is the 'imitation' of the child by his caregivers (Pawlby, 1977). This is a ubiquitous phenomenon. Mothers (and fathers) hear their child making sounds and later protowords and words, and they reflect these back to the child not as exact, mimicked renditions but in well-formed syllables of L1. In other words, they 'reformulate' the child's output, based on their judgment of what it sounded most similar to in their phonological system. If the child (i) knows what he did (which he often will, whenever he was using a so-called 'vocal motor scheme'[2] that he has automatised), and (ii) knows that he was being imitated (which an infant recognises from an early age), then he can conclude that at least as far as his mother is concerned what he did and what she said are equivalent. When he explodes his lips apart with his tongue high in his mouth and his vocal folds vibrating, his mother takes her turn in the game by saying [bi:] back to him.

This is enough for him to now enter into learning the pronunciation of words. When he recognises an element in a word his mother says for which he has a corresponding production routine he can deploy this in attempting to say the word.

Obviously there is now a long story of further development, but the key interactions for learning speech sounds are, as described, the episodes where his mother acts as a vocal mirror for articulatory routines that he has developed for himself. In these episodes, she makes a judgment of similarity between his and her output, but he need not. He only has to realise that she has made that judgment, and that he can make use of the equivalence/ correspondence that she has demonstrated between his action and her sound.

Is all this true?

I imagine every reader is now asking a very good question: is all this true? Do children really not learn either 'timing' phenomena or speech sounds by imitation? I think they probably don't, but I haven't carried out a test that would decisively tell between the conventional accounts and the new ones I am proposing. So my view is based on the relative plausibility of each account from various perspectives: psychological, the data from developmental studies, anomalies (like PFC in the speech of young Swedish children), and so on. I discuss all this in detail in the thesis, but here I would just like to make one important point that is not generally appreciated.

There is no evidence whatsoever (not a scrap!) to support the conventional accounts. They are, and have only ever been, no more than assumptions. Similar assumptions have, of course, been wrong in the past.

With respect to the timing of speech, there is even a name for the mistake that I think we have made: William James's "psychologist's fallacy", which Ladefoged (1984:92) summarised as "the notion that because an act can be described in a given way [by an observer] that it is necessarily structured in that way [for performance]."

With respect to speech sounds, while it may appear to be common sense that children develop these by imitating what they hear, it is also common sense that the earth is at rest and that the sun travels across the sky. Now, an analogy with pre-Copernican thought is rather grandiose for the subject we are considering, but I hope it may help to break the hold that 'imitation' has on our

[2] McCune and Vihman (2001) define vocal motor schemes as, "… generalized action patterns that yield consistent phonetic forms," and, "[a]ny consistently occurring phonetic pattern developed, in theory, by repeated and regularized child action.

imaginations. For not only is the assumption of an imitative mechanism for speech sound development unsupported by any evidence, but it also leads to numerous theoretical problems and thus a current view of speech which is not dissimilar in its horrible complexity to the Ptolemaic system of epicycles and eccentrics. (A view of speech, in fact, which is well characterised by the despairing phrase "speech is special" to 'explain' the psychological perversity of what we presently imagine to be true.)

Learning and teaching the 'timing' phenomena of English

Let us assume for now that my proposals are well founded. What implications do they have for teaching?

With respect to the timing of speech, it is easy to recommend one of the things we should stop doing. We should stop teaching 'rhythm' *per se*, i.e. as if native speakers end up with the temporal organisation of English because they are adjusting the timing of segments to achieve a 'stress-timed' rhythmic effect. Instead, we should concentrate on the various non-temporal sub-processes that contribute to this overall percept, and then let the result be what it is. (This is not a new suggestion, but the explanation I have put forward for the phenomenon of foot-level shortening strengthens the case made by others. Without such an account it is hard to see how we can avoid bringing timing into our teaching of prosody.)

In Messum (2002), I described Catford's proposal (Catford, 1977; 1985) on one sub-process of 'rhythm': the reduction of full vowels to schwa, etc. He argued that for pedagogical purposes weak vowels are better viewed as open transitions between flanking consonants than as proper vowels. I have found this to be very helpful.

Here, I would propose that we become more precise about our presentation of another such sub-process: the production of routine sentence stress. We already ask students with non-stress-accent first languages to make a stressed syllable prominent by making more 'effort' on it. We illustrate this with forceful arm gestures etc, but it is often operationalised by our students as just increased loudness. Unfortunately, loudness can be increased by a laryngeal adjustment alone, with no, or only a minimal, increase in respiratory drive. Because a French, Thai or Japanese speaker will feel it is rather odd to make several transient increases in his respiratory drive during a breath group, I suspect that when doing an exercise in class he will satisfy the teacher by just making syllables louder in the least effortful way, through laryngeal adjustment.

This has two disadvantages: (1) it does not anchor English stress to any motor activity that can be felt directly, and (2) it does not affect the aerodynamic aspects of his speech in any way that could result in natural foot level shortening as a result of the allocation of limited aerodynamic resource (or result in other, similar 'timing' adjustments).

On the first of these, I suspect that by copying loudness as a purely acoustic effect, our student will have given himself a way of speaking English that cannot be straightforwardly automatised. That is, while he is in class and concentrating on saying words with an authentic acoustic pattern of stress he may succeed, but when he tries to speak without this close attention being paid to how he achieves prominence then he will revert to whatever mechanisms of prominence his L1 uses. It will be just too uncomfortable and too unnatural to routinely make syllables louder in the way he has practised.

The solution, I suspect, is to make sure that our students learn to stress syllables with perceptible motor activity that increases respiratory drive. That activity should probably be action of the abdominal muscles, which can be both consciously directed and felt. This is not to say that native speakers employ this mechanism in their relaxed conversational style. But in the same way that we initially learn oral articulations through hyper-articulation this is probably the best entry point into pulsatile expiratory speech breathing.

The abdominal respiratory muscles run vertically, obliquely and in a ring around the belly. They drive air out of the system by compressing the contents of the abdomen, which pushes the diaphragm upwards, which reduces the volume of the lungs. The process is a little like squeezing to get toothpaste out of a tube.

It is quite straightforward to become aware of their activity, and in speech and language therapy an approach called the Accent Method (Thyme-Frokjær and Frokjær-Jensen, 2001) has simple techniques for improving the way that speakers use these muscles which seem easy to adapt for TEFL. If we are adamant that our students should increase their respiratory drive in this way on stressed syllables then they will certainly sound more natural – with increased loudness and length coming automatically – and they will have a tangible action associated with stress which they can carry out of the classroom into their free speech.

On the second disadvantage of our present approach that I mentioned, it is less certain that a modification to his speech breathing will help our student to replicate FLS, the lengths of tense and lax vowels, etc. The event that precipitates these adaptations by English speaking children is their adoption of stress-accent as the routine means by which syllables will be stressed. However they do this with a speech production system whose mechanics and aerodynamics constrain child speech in a way that the adult counterparts do not.

Time will tell on this. It is encouraging, though, that some non-native speakers do master English prosody even as late learners of the language. And we can be confident that working on use of the abdominal musculature for stress-accent will at the very least make it easier for students to aspirate their plosives and produce a good /h/ sound without disrupting the natural flow of their speech.

I did not do enough teaching over the years it took me to produce my thesis. But the little I did which incorporated these ideas has been encouraging. I used nursery rhymes and well known sections of Shakespeare as texts. The former have a proven efficacy in forcing young speakers to reduce vowels, produce feet with single pulses, etc, and the latter provide a challenge to grown-ups to do the same.

Learning and teaching speech sounds

We are very fortunate that Caleb Gattegno demonstrated how to teach speech sounds without involving "listen and repeat" or any of its variants. (His work, in fact, was the starting point and inspiration for my thesis.)

Gattegno called his approach to teaching foreign languages the 'Silent Way' (SW) (Gattegno, 1962; 1976). The name reflects the 'silence' of the teacher, who doesn't model language and doesn't give answers that students can find for themselves (Young, 2000). So being 'silent' is not the same as being mute; but the fact that a teacher can teach speech sounds, in particular, without modelling them is certainly a striking aspect of the approach.

For anyone who is unfamiliar with the SW, Gattegno devised a series of charts that freed him from the need to provide pronunciation models. The first is a sound-colour chart, a collection of coloured rectangles that are roughly equivalent to the phonemes of a language. There is an example of how one can be laid out for English in Messum (2002). The colour coding system is used for the vocabulary and sound/spellings charts that are introduced later.

The sound-colour chart enables a teacher to establish correspondences between syllables of English (indicated by one or more taps of a pointer on the rectangles) and the syllables produced by the class, without him having to say anything himself. As students become familiar with the chart, long utterances can be constructed this way, and some aspects of the prosody of the language profitably worked on before students are aware of the meaning of what they are saying. However, here I will confine myself to discussing the teaching of speech sound qualities, and in all its essentials the way that Gattegno did this reproduces the mechanism used by young children that I described earlier.

He encouraged students to experiment with their own sound production apparatus in the absence of an acoustic model, shaping the vocal motor schemes they developed by indicating where they were not acceptable for the language being learnt and where they were (for the moment, anyway). The correspondence that a caregiver demonstrates through her reformulation of the young child's utterance was established not with the teacher's voice but with the rectangles on the chart.

This begs a question. Why not faithfully reproduce the LI interaction in all its details? There are several reasons for this. One is that many students would find the temptation to mimic the teacher overwhelming. Among the problems that mimicry brings, a fuller account than I have space for here would discuss (i) tissue conducted sound, (ii) the differences between the teacher's and students' voices, and (iii) the students' lack of criteria for judging similarity within the phonology of the language they are learning. However, the most important problem is that even successful mimicry does not, in itself, lead to the learning of an automatised vocal motor scheme for speech.

To see why, imagine that a student is learning to produce a vowel that is not in his existing repertoire. This does not mean that he cannot perform something like the vowel in question, given a model to mimic. Mimicry means the matching of two events on a sensory level, and our powers to do this are quite distinct from the motor routines we develop for automatic speech production. With an auditory (sensory) image to work from we can impersonate an L2 speaker without having any ability to speak L2.

Such an auditory image can be formed in short-term memory from a model that I have just been presented with, or it may be something that I can evoke from presentations in the past. I may be able to match it using old learnings about my voice, but in doing so I am not learning anything new, and, to the extent that I continue to attend to the matching of two sensory experiences, my attention will not be on what I need to be aware of in order to develop a vocal motor scheme that will function in the absence of an auditory image. I need to develop a motor automatism of the latter kind to become a speaker of L2 or L1 (meaning that I will not have to evoke every word into conscious awareness before I can say it).

Gattegno recognised that mimicry was a dead end in learning to pronounce, and the teacher's silence forces students to abandon any reliance on it. Instead, they will attend to what they do (at an articulatory level) and what the results are. These results are evaluated linguistically by an expert listener (the teacher) who can be sensitive to how this particular student, with his or her unique voice, should sound when speaking L2.

It does take a leap of faith to teach your first lesson this way. It helps, of course, if you have been a student in a class with an

experienced SW teacher first. It also helps if you have some understanding of articulatory phonetics, so that you can help students with suggestions that may assist their trial and error process. All teachers should have this, of course, and any reader of Speak Out! will certainly have sufficient expertise. But once you have taught a few lessons this way, my prediction is that you will be 'hooked'. The absence of the teacher's voice in the learning space is liberating for the students and for the teacher. Real progress with pronunciation problems can be made.

Conclusion

I started by suggesting that if children don't learn to pronounce by imitation then we should reconsider our teaching practice. In fact, the techniques I recommended have already been developed by others and shown to be effective. It would be beneficial to pronunciation teaching if they were adopted more widely whether or not my accounts of child learning are vindicated. However, more people will certainly be persuaded of this if we discover that children's pronunciation is indeed shaped by the aerodynamics of speech in a child-size body and by the imitative games children play with their caregivers.

Happily, there are good reasons to believe that this is the case. We can therefore expect better results in our work in the future if we are prepared to modify in some ways our approach to teaching pronunciation.

References

Buder, E. H. & Stoel-Gammon, C. (2002). American and Swedish children's acquisition of vowel duration: Effects of vowel identity and final stop voicing. *JASA*, 111/4, 1854–1864.

Catford, J. C. (1977). *Fundamental Problems in Phonetics.* Edinburgh University Press.

Catford, J. C. (1985). 'Rest' and 'open transition' in a systemic phonology of English. In *Systemic Perspectives on Discourse Vol 1*. W. S. Greaves & J. D. Benson (Eds.), Ablex: Normal NJ.

Crystal, T. H. & House, A. S. (1988). The duration of American-English vowels: an overview. *Journal of Phonetics*, 16, 263-284.

Dauer, R. (1983). Stress-timing and syllable-timing reanalysed. *Journal of Phonetics*, 11, 51–62.

de Jong, K. J. (1991). An articulatory study of vowel duration changes in English. *Phonetica*, 48, 1–18.

Gattegno, C. (1962). *Teaching Foreign Languages in Schools: the Silent Way.* Reading: Educational Explorers.

Gattegno, C. (1976). *The Common Sense of Teaching Foreign Languages.* New York: Educational Solutions.

McCune, L. & Vihman, M. M. (2001). Early phonetic and lexical development: a productivity approach. *Journal of Speech, Language and Hearing Research*, 44, 670–684.

Ladefoged, P. (1984), 'Out of chaos comes order'; physical, biological, and structural patterns in phonetics. In *Proceedings of the Xth ICPhS.* A. Cohen & M. P. R. van den Broecke (Eds.), Dordrecht: Foris.

Messum, P. R. (2002). Learning and teaching vowels. *Speak Out! (Whitstable, IATEFL)* 29, 9–27.

Messum, P. R. (2007a). *The role of imitation in learning to pronounce.* PhD thesis, London University.

Messum, P. R. (2007b). How children learn to pronounce: not by imitation but by their mothers' vocal mirroring. Unpublished MS submitted to 15th ICPhS, Saarbrucken.

Messum, P. R. (2007c). Mirroring, not imitation, for the early learning of L1 pronunciation. *JASA*, 122/5.2, 2997.

Messum, P. R. (2008). Embodiment, not imitation, leads to the replication of timing phenomena. *JASA*, 123/5.2, 3329.

Naeser, M. A. (1970). *The American child's acquisition of differential vowel duration.* Center for Cognitive Learning, Technical Report No. 144. University of Wisconsin.

Pawlby, S. J. (1977). Imitative interaction. In *Studies in Mother-Infant Interaction*. H. R. Schaffer (Ed.). London: Academic Press.

Thyme-Frøkjær, K. & Frøkjær-Jensen, B. (2001). *The Accent Method.* Bicester: Speechmark.

Young, R. (2000). The Silent Way. In *Routledge Encyclopedia of Language Teaching and Learning*. M. Byram (Ed.). London: Routledge.

Piers Messum is a freelance teacher from London. He is recovering from taking 11 years over doctoral research that he thought would take 6 months. He has taught in France, Japan and the UK.

Contact him at p.messum@gmail.com.