Speak Out!, IATEFL (Whitstable, Kent), 2004 32:12-23

Autonomy, as soon as possible

Piers Messum
p.messum@ucl.ac.uk

I have taken my title from Einstein's dictum that, "a scientific theory should be as simple as possible, but no simpler." Similarly, it is in our students' interests for them to be independent and autonomous learners as soon as possible, but no sooner. This article is about how we and they might make the judgment of when independent work on pronunciation is appropriate and possible.

To explain my thoughts on this, I will report the highlights of the research I have been carrying out into (1) how children learn to talk and (2) why, therefore, the phonetics/phonology of English in mature speakers is the way that it is. I will then relate this to L2 learners, in order to ask what we can realistically expect students to do on their own, with and without help from technology, and what activities will need a teacher or some other source of sensitive and informed feedback.

Some of my own experiences as an independent learner of pronunciation are still very vivid in my memory, and I think that I am now in a position to explain why the good experiences were good and the bad ones were bad. Among the good ones were the following:

- On several occasions when learning Japanese I found myself babbling away (to myself) all the way home on the bus.
- A song I learnt in a French class many years ago has stayed with me, and I sing it to myself whenever I want to feel what I think it is to be a French speaker (with all that lip movement and nasalization).
- When I first noticed that French children seemed to be hyperarticulating their speech this set me off on hours of practice on my own to see if I could do the same and with what results. The discovery that French is written phonemically with no gaps between the words set off a similar burst of practice, this time on liaisons.

On the negative side, though, I have invigilated many language laboratory sessions when the purpose was improvement of production skills which both the students and I found frustrating and ineffective. The pedagogical paradigm underlying this work was the idea that improving the students' listening would improve their pronunciation, so sooner or later students were asked to imitate a model provided on the tapes. In this way the technology was only following what happens in many pronunciation classes, but in such a concentrated form the shortcomings of this approach were particularly apparent.

It seems to me that I can now explain these and other experiences in pronunciation learning through the framework I develop below. I will start, though, not with L2 learners but by looking at how children learn to talk, with 'talk' used here not in its colloquial sense (where it is often synonymous with 'speak') but with a restricted meaning encompassing the phonetic/phonological part of speech but not the cognitive or linguistics ones.

Do children learn to talk by imitation?

There is a widespread belief that children learn to talk by imitation. I will come back later to exactly what this might mean, but the fact that children faithfully replicate (or 'match') the important phonetic/phonological features of adult speech has led most people to this conclusion. However, to the best of my knowledge the notion has never been tested, and the few serious attempts to examine it that I have read seem more puzzled than convinced by it[1]. It seems to be an assumption made in the absence of any other hypothesis rather than a truth that has been demonstrated.

It should be said that there have been good reasons for believing an imitative account: there are phonetic phenomena which, it has seemed, could only be replicated by imitation and I will deal with some of these later. However, in the studies of imitation in its many manifestations one recurring theme is that behaviour which appears at first sight to be imitative often turns out to be something else on closer examination. A stranger to tennis, for example, might observe the synchronised sideways movements of 20,000 heads around the Centre Court at Wimbledon and imagine that a newcomer, who has quickly started to do the same as everyone around, has picked this behaviour up by imitation. Only when the observer becomes aware of the action on court would he understand that the matching was the result of a common goal and similar bodies being used to achieve it, rather than any copying process.

To develop an alternative to the imitative account of children learning to talk, we need to be clear about some of the phenomena that are grouped under the label 'imitation' and to examine how social learning (learning from others) takes place. In Messum (2002) I wrote about these things as part of an article for Speak Out! on the learning and teaching of vowels, and I would now like to take the opportunity to revise and update what I said then.

Types of 'Imitation'

In that previous article I struggled with the way that 'imitation' is often used on the one hand whenever there seems to be any connection between one person's behaviour and the behaviour of someone he has observed, but on the other in a more precise fashion when the explicit copying of that behaviour is being asserted. I ended up using 'mimicry' to describe

[1] With a notable exception in Gattegno, who was neither puzzled nor convinced, but instead wholeheartedly rejected the idea He expressed this in a forthright manner in an oral address to the Association of Teachers of Mathematics just prior to his death:
"Have you ever noticed that children learn to speak their mother tongue by themselves? And that you are evading questions in saying, "They do it by imitation." "By imitation," indeed. The greatest nonsense I ever heard, and everybody repeats it. It's absolutely wrong. No-one can learn to speak the mother tongue by imitation. So, you have to ask the question: how did we - because we were babies - how did we learn our mother tongue? What sort of powers of the mind did we have to sort these things out by ourselves?" (Gattegno 1989)
He gave reasons for this viewpoint and an account of how babies and young children do learn to talk (and then speak) in 'The Universe of Babies' (1973).

the latter process, but a better solution (which doesn't distort the meaning of 'mimicry' in the way I did) is described by Call and Carpenter (2002).

They point out that in situations where the behaviour of a demonstrator and an observer match, the demonstrator will have supplied three independent streams of information: about his goals, his actions and the results (i.e. how the environment is changed by the actions). They explain that depending upon whether the observer adopts the goal(s), copies the actions or achieves the same results, we should describe his behaviour in different ways, as summarised in the table below.

	Goals	Actions	Results
Mimicry		✓	n/r
Imitation	✓	✓	✓
Emulation			✓
Goal Emulation	✓		✓

For humans (as opposed to animals) it is probably not important to make a fine distinction between emulation and goal emulation. All the experimental work recently done with children suggests that in humans goal emulation is the norm, so I shall use plain 'emulation' to mean goal emulation from now on.

Thus the terms of interest to us could be defined as follows:

Mimicry, where the observer either does not understand or does not adopt the demonstrator's goal, but does copy his actions. Whether he reproduces the result or not (i.e. somebody responds to or acts upon his message in the case of voice mimicry) is not relevant.

Imitation, where the observer understands and adopts the demonstrator's goal, copies his actions, and reproduces the result. If the observer fails to achieve the result, then we would describe this as failed imitation.

(Note that the 'copying' of actions is highly problematic in the case of speech – are the actions the demonstrator's articulatory gestures, the sounds he makes, or what? – but this issue is not important for my argument here.)

(Goal) emulation, where the observer understands and adopts the demonstrator's goal and reproduces the result, but doesn't copy his actions (intentionally).

An example of emulation rather than imitation is the way that George Bush Jnr emulated his father's achievement in becoming President of the USA. His goals and the result may have been the same as his father's, but his route to the top was his own, via governorship of Texas rather than being a senator, Director of the CIA, etc.

Learning "by imitation"

One important point to appreciate, however, is that the phrase "learning by imitation" (or "by emulation", for that matter) obscures the real processes involved. We don't learn by imitation: imitation is something we do when we have already learnt. Roslyn Young makes this point well:

> When I was a child, I learnt to walk a tightrope. So if I installed a rope between this rooftop and that one, and told you, "Now off we go. Just do like me." would you try? Of course not, because you know as well as I do that you can't imitate me in this. You have to develop the sensitivity to your centre of gravity and all the other technical skills and the muscular power in your feet and in your abdomen which will allow you to do it. This is an extreme example, but a little thought shows that in all circumstances, without exception, it is only possible to imitate what one can already do. If I don't already possess the gesture I can't imitate someone else doing it.
> If imitation were part of the learning process we could all be champions in any discipline we wanted to. Just watch and do.
> Imitation exists of course, but when someone is imitating they are not faced with the unknown. They are using skills which they already possess. (Young 2000)

If there is any bridging of the gap needed between what an observer can already do and what he needs in order to imitate, then the learning required takes place in the normal way prior to the act of imitation: through awareness, practice etc.

As part of these preparatory remarks, I would also like to introduce a distinction made between so-called 'reinforcement' learning and 'supervised' learning (Wolpert et al. 2001). The difference relates to how the learner comes to adjust his performance. In reinforcement learning the environment provides just binary feedback: performance is acceptable or unacceptable. In supervised learning the environment provides a target, and performance can be measured by the discrepancy or error between the actual and this desired output.

When we say that the target has been internalised by the learner, we mean that he has his own criteria for assessing his performance, and his learning can then be 'self-supervised'. The importance of these distinctions will become apparent shortly.

Acquiring skills in the real world

Having set up a contrast between imitation and emulation, however, we have to recognise that when we learn from others in multiple cycles of improvement our approach may vary over time. Thus Morrison (2002:115) points out (using 'imitation' in this quotation in the wider, colloquial sense rather than the technical one I defined above) that,

> When humans imitate, we neither emulate another person's goal irrespective of the means to achieve it, nor do we mindlessly regurgitate action sequences without regard

> to what they accomplish. Rather, means and ends go hand in hand (so to speak) when we imitate an action.

Similarly Call and Carpenter (2002:219) explain that,

> … it is a common experience among adult humans to watch someone achieve some result (e.g., with a new tool, or when learning to play a new sport or musical instrument) and then to attempt to reproduce that result oneself. If one's first attempt is unsuccessful, during the next demonstration one might pay more attention to the demonstrator's actions than to the end result.

So the cycles of attempts at a new skill that take place in normal learning are a mixture of emulation and imitation (and sometimes, perhaps, just mimicry). Depending on what we perceive to be the best next step towards our goal we may, for example, pay attention to discovering the contingencies of a situation for ourselves, or pay attention to what someone else has discovered and attempt to copy their solution to a problem.

However, viewing this process with adult eyes we must be careful not to assume that insightful perception of the demonstrator's actions is a given. In learning something new, perception may need education as much as our motor systems. I would approach the new skill (to me) of learning the violin with a lifetime of watching other people's actions and listening to music behind me, so my perception would immediately permit some basic degree of imitation (to the extent, that is, that my motor system allowed). But I have never listened to the violin for anything other than its effect on me. To learn something from another violinist's playing I would have to learn to listen with sensitivities I suspect I am not even aware of at present. Jenkins (1980) gives further examples of this, in an article about research in child phonology:

> When one wants to train an adult to draw, one begins by "training the eye." That is, artists believe that adult non-artists, far from having perfect perception, have uneducated and undeveloped perceptions. Art teachers have a host of exercises devoted to training people to look at the world in such a way that they see those aspects of things that the artist believes to be important. In another example, there is evidence that the perceptions of chess experts are not at all like the perceptions of novices. (p. 225)

With respect to children learning to talk, their perception will have been trained to some extent for understanding what is being said to them, but there is no reason for us to assume that it will be sensitive to fine phonetic detail such as timing distinctions that are not linguistically important, or, indeed, to what their own voices sound like in relation to the material they perceive from others[2]. If this is the case then self-supervised learning (and

[2] There are, in fact, two issues here. Firstly, what a child's voice sounds like in relation to others, given that a child (i) has a voice with different pitch and loudness ranges, (ii) is not capable of reproducing an adult formant space, (iii) speaks at a different rate, and so on. Secondly, that the child is anyway not 'hearing' the signals in a comparable way: the 'sound' of his own voice will be significantly altered by the part of the signal transmitted through his body tissue.

hence imitation) will not be available to them. They will only be able to make use of feedback from the environment along the lines of being understood or not being understood to guide their predominantly reinforcement learning.

This will change over time, of course. Voice imitation exists, and children come to be able to imitate others. But at the start there is no reason to believe that it is available to the young learner[3].

Learning phonetic phenomena

We can now return to how children learn to talk. The features of spoken language that will be replicated by an English-speaking child include the following:

1. Speech sounds: vowels, consonants, etc.
2. So-called language 'universals': effects like declination and the shortening of vowels before fortis consonants (i.e. realising the vowel in *cat* as shorter than that in *cad)*[4].
3. Language specifics: for example, the 'rhythm' of English, or the different lengths of tense/lax vowels, as in words like *heat* and *hit*[5].

Generally speaking, these features are believed to be copied by the child from models in his linguistic environment. Some doubts are expressed from time to time about speech sounds being acquired this way, and physiological arguments are sometimes advanced to explain language universals, however the third category of features seems to provide incontrovertible evidence of imitation, for how else could language-specific phenomena be replicated? Having attributed to a child the will and ability to use imitation for this category, most speech researchers then extend them to his acquisition of the other features of speech.

In contrast, my research shows how imitation is probably not the mechanism for replication of features in any of the three categories above. With respect to language-specific features (in the crucial third category), I show that these may not be primary phenomena, but may instead be epiphenomenal – the surface manifestations of a more primitive set of speech adjustments which are motivated by the need to rebalance the aerodynamics of speech when, as in the case of English, this has been unbalanced by the adoption of stress-accent to signal syllabic prominence.

I am also not sure about the extent to which we 'listen' to our own voices, anyway. We certainly monitor our output, but this may be via a forward model of the expected sensory results of motor commands, rather than via an aural pathway.

[3] Unless 'speech is special', as some researchers have asserted. Thus Lieberman (1980:137) hypothesized the presence of an innate perceptual mechanism that normalizes speech, in order to explain how vowels could be imitated.
In another field, the imitation of facial gestures, Meltzoff and Moore (e.g. 1997) have granted similarly 'special' powers to babies, but there are good arguments against their interpretation of the data (see e.g. Jones 1996).

[4] These effect may be suppressed in a language like Swedish where vowel length is phonemic.

[5] In fact, and importantly, these length differences only occur when vowels are in prominent (stressed) positions.

I give details of how I think this happens in my thesis[6], but, in brief, with respect to the 'rhythm' of English I propose that rhythmic clipping/foot level shortening (the process by which syllables get more compressed as the number of syllables in a foot grows – e.g. in | *one and then* | *two and then* | *three* | vs. | *one* | *two* | *three* |) is the result of the need to distribute a given amount of respiratory system energy over a variable number of segments. One result is adjustment in the timing of all the segments in the foot, but this is not because the speaker is attempting to speak rhythmically. The changes are by-products of another process.

Similarly, the length increase in tense vowels in prominent positions is motivated by the need for a young (and small) learner to avoid turbulent airflow at the point of maximum constriction in his mouth. (Turbulence would create a sound quality which is unacceptable for a vowel.) So he reduces airflow by increasing resistance at his glottis, and the sound is prolonged so that he uses up the full quantum of respiratory system energy he has applied.

Similar arguments, based on physiology and aerodynamics, explain the replication of features in the second category, above, that of language universals.

With respect to speech sounds (in the first category) I propose that a normal cycle of learning will explain their acquisition adequately, without the need for us to invoke special, innate mechanisms. As I have already indicated, one crucial aspect of this is that the child's perception does not allow him at the start to imitate, for example, vowels. However he can certainly attempt to emulate the extraordinary effects that he sees others achieving with their voices, learning by reinforcement from his interlocutors' acceptance or non-acceptance of his attempts at speech.

The possibility that a process like this may occur has now been tested in a limited way with 8 month olds, with positive results (Goldstein et al. 2003). A similar process seems to explain the acquisition of song by some birds (Smith et al. 2000). In the field of language teaching we have a 'demonstration' proof of the effectiveness of learning this way in the success of Gattegno's Silent Way methodology.

I discussed why learning without a model and without imitation is successful in my earlier article. To summarise what the learning process is, though, I would like to quote again Roslyn Young's description:

> First of all, learning a new sound requires that the student realize that there is in fact a new sound to learn. He can then try to create the sound. In this case he is dealing with two independent but closely related systems, the mouth and the ear. Only one of these systems, the mouth, can be controlled voluntarily. All the muscles of the ear are involuntary muscles. The student can only modify the voluntary system. With his mouth he produces a sound which he guesses might be as close as possible to the sound he is aiming for. He hears the sound with his ears. Since he produced it with his own

[6] To be submitted, I hope, very shortly. I will be happy to send a copy to anyone interested in properly understanding the mechanisms I only sketch here.

> mouth, he knows that, muscularly speaking, his mouth was used in a new or special way and consequently he knows he should listen for a sound which is different from what he usually hears. He can probably predict at least to some extent in what ways the sound will be different from what he usually produces. He speaks here with the deliberate intention of hearing something unusual and he listens to the result with the specific intention of hearing this unusual sound he has produced, creating a double feedback loop. He has feedback from his mouth telling him what it is doing and his ears give him feedback about what changes they detect as a result. Gattegno proposes that this is the process we all use to learn to produce new sounds.
>
> Once the student has managed to produce the sound to his satisfaction, he must practise it in a wide variety of different situations and contexts until he is completely at ease with the sound. He then reaches a stage where the sound has become completely automatised and the learning process for that particular sound is over. (Young 1995)

Generalising this to other features of speech (for example to learning the use of stress-accent in English to make syllables prominent) we could describe four stages of learning:

1. Awareness of a feature of speech as profitable to adopt
2. Rudimentary attempts by the learner, making use of previous learning (babbling for young learners, L1 for older ones) but using production- rather than perception-based criteria for control. Predominantly reinforcement learning, i.e. more emulation than imitation.
3. The gradual education of the learner's perception, giving increasing potential for self-supervised learning and thus, in due course, the possibility of imitation, mimicry, conscious stylistic change, etc.
4. Transfer of control from production- to perception-based criteria[7].

This is a normal cycle of learning, similar to the process by which children might learn to catch a ball or build a tower with blocks. It does not require any special, innate abilities, specific to speech. Use of this mechanism, plus accommodations to the aerodynamics of speech in a small body, can explain the appearance of all the speech features that a child replicates in learning to talk.

Implications for pronunciation teaching, and independent learning

If the process above is what children go through then this gives pronunciation teachers both opportunities and problems.

On the positive side, we can better explain and justify the approach to the learning of speech sounds proposed by Gattegno in the Silent Way[8]. I have only rarely had a negative reaction

[7] Or, perhaps, a forward model of the type I mentioned earlier.

[8] Perhaps I should note again that the essence of the Silent Way is the subordination of teaching to learning, not the materials, technologies or techniques often associated with it in the public mind. The particular techniques Gattegno devised – use of wall charts, finger correction, etc - can be used in many ways where there is no

from a student when I don't model sounds to a class and instead ask students to find them by experimentation (indeed students seem to hugely appreciate the chance to speak in an environment not dominated by the teacher's single, correct model). Any negativity I have experienced has always quickly gone away when we start working. However, before having the practical experience of learning this way students are certainly surprised to hear about this approach, and teachers seem to be wary about adopting it. (This may be for reasons other than scepticism: for example, a nervousness about how to go about giving feedback on pronunciation without modelling the 'correct' answer.)

It is not a given, of course, that L2 learners need to retrace the path that L1 learners take with respect to speech sounds. Clearly the former bring vastly more knowledge and articulatory skill to their learning of languages subsequent to their first one. However if we trace where their awareness needs to be directed during the learning of new sounds then it seems that they must go through a similar process and that they often start with similar perceptual insensitivity (an inability to perceive the distinguishing features of some new sounds).

This being so, the initial learning of L2 students can only be undertaken by emulation rather than imitation, and they will require feedback from the environment that efficiently helps them to refine their attempts at acceptability. As far as I am aware, no practical/affordable technology exists to give them that feedback apart from the language teacher. For this aspect of pronunciation, then, independent work and early autonomy should be approached with caution. However, when students have acquired some criteria for the way that speech sounds should be produced there will be unlimited scope for them to practise applying these to the sounds they produce in new words, when speaking louder or softer, at different rates, and so on. It seems to me that my babbling in Japanese on the bus and the other episodes of independent learning I described earlier were triggered by my development of criteria about how the languages I was learning should be produced, and sustained by my excitement in applying these new discoveries.

Apart from speech sounds, stress-accent (making syllables more prominent by putting more respiratory system and general effort into producing them) is also something that a child seems to learn through cycles of mainly emulation and then later, perhaps, some imitation. Exercises where we encourage learners with native languages that do not employ this device to put more physical effort into stressed syllables would therefore seem to be the right approach to teaching this[9].

---

subordination at all, in which case they will be no more effective than any other teaching-led activity, even if their appearance is superficially faithful to SW practice.
Also, even when used correctly they are only the techniques which he devised, in the face of the particular learners he encountered; we need to and can devise other solutions to our own problems if techniques that he and others created are inappropriate. These will be 'Silent Way' as much as anything Gattegno did if they are based on a correct understanding of how language is learnt and if they respect the pedagogical principles that underlie the notion of the subordination of teaching to learning.

[9] In languages such as French or Japanese, syllables are signalled with routine prominence by length changes and pitch movement, but not usually by differences in loudness. Length changes can be achieved with or without involvement of the respiratory system; in the case of English, greater effort will naturally lead to

However, the account of children learning to talk I have given suggests that many features of English pronunciation – less significant ones but nevertheless highly characteristic – emerge as a result of processes that are motivated by aerodynamics etc. It is an open question at this stage whether the conditions that precipitate these changes only apply in child-size speech production systems and not adult-size ones: the speech breathing mechanics and vocal tract aerodynamics of adults are very different from those of children.

If this is the way that 'rhythmic'-type effects and tense/lax vowel characteristics emerge, then it explains the resistance of many students to attempts we make to teach these things explicitly. Even in the case of students who enthusiastically undertake exercises along these lines, my suspicion is that there is very much less carryover here from the classroom into normal speech than we would like.

Nevertheless, there are a number of people with L1 languages very different from English who come to pronounce the language very well. This suggests that the aerodynamic processes that affect children can also affect adults, at least to some extent. If so, then there may well be ways to teach which will lead to the emergence of the phenomena in the second and third categories I gave earlier.

Certainly some sub-skills that contribute to these phenomena can be taught. With respect to rhythm, for example, I am suggesting that the 'rhythmic'-clipping (described above) that leads to an appearance of stress-timing is not motivated by rhythmic concerns at all, but rather by the allocation of limited aerodynamic resources. But this process demands a high degree of upper articulatory prowess and an understanding about what can be sacrificed and what needs to be preserved in English when its syllables have to be compressed. This allows the speaker to produce each foot within the constraint imposed by his speech breathing mechanism.

Now, one way that English speaking children learn these skills (so-called 'vowel reduction', elision, assimilation, etc) is through nursery rhymes, whose genuinely rhythmic nature encourages speakers to warp their natural production of segments to fit in with its demands. It would seem that practice with similar material should help L2 learners, as long as we are clear that this is to help them with foundational skills, not that normal English is spoken 'rhythmically'.

Of course, nursery rhymes are marvellously memorable and may therefore be a tool which facilitates independent work by the learner in the same way as the French song that has stood by me over the years.

I described some other ideas for a 'pedagogical phonetics' in Messum (2002). For example, teaching word medial reduced vowels as 'open transitions' (ways of getting from one

---

increased length and/or loudness (in a trading relationship between the two), while in French the length is probably controlled directly.

completed consonant to the next, rather than real vowels), should also facilitate student production of natural English prosody.

Conclusions

I started by voicing my concerns about the possibility of independent work by students when it comes to pronunciation. I hope that it is now clear that once a learner has developed criteria for what is acceptable and what unacceptable in his production, then he will be in a position where self-supervised learning is, in fact, possible.

However, where criteria are not in place – and this is true for many supposedly advanced learners of English with respect to quite basic aspects of their pronunciation – then independent work is probably not possible. The student must be encouraged and allowed to experiment with new ways of producing the language, with feedback from a sensitive teacher as to the acceptability of these attempts. Any work based around an imitative model – whether in class or a language laboratory - is likely to fail, at this stage. (Except, perhaps, to the extent that the student subverts the process by going beyond what the teacher has actually demanded (Messum 2002).)

We have a model for how to approach pronunciation teaching in Gattegno's Silent Way. However, techniques other than his can and must be devised, and there are many examples of this which are consistent with his pedagogical philosophy[10]. If my proposals as to how features of English such as (so-called) 'stress-timing', tense and lax vowel characteristics, patterns of aspiration and so on are correct, then we need to rethink how these are taught, but if we can reproduce to any extent the way they emerge in children then students will have the bonus of them appearing without the effort that is now put into their mastery as explicit processes of the language. To that extent, our learners will certainly be autonomous in their pronunciation of English sooner than we can help them to be now.

References

Call, J. & Carpenter, M. 2002, "Three sources of information in social learning," in *Imitation in Animals and Artifacts*, K. Dautenhahn & C. L. Nehaniv, eds., MIT Press, Cambridge, MA, pp. 211-228.
Gattegno, C. 1973, *The Universe of Babies* Educational Solutions, New York.
Gattegno, C. 1989, "A farewell address," in *A Gattegno anthology: Selected articles by Caleb Gattegno reprinted from 'Mathematics Teaching'*, L. Brown, D. Hewitt, & D. Tahta, eds., Association of Teachers of Mathematics, Derby, UK, p. 13.
Goldstein, M. H., King, A. P., & West, M. J. 2003, "Social interaction shapes babbling: testing parallels between birdsong and speech", *Proc Natl Acad Sci USA*, vol. 100, no. 13, pp. 8030-8035.

[10] Adrian Underhill's 'Sound Foundations' (1994) contains many such exercises. A few of these rely on ideas drawn from phonetic theory that I am now arguing against, but his general approach – working with presence and awareness, developing alternatives to correction, and so on – seems absolutely right.

Jenkins, J. J. 1980, "Research in child phonology: comments, criticism and advice," in *Child Phonology. Volume 2, Perception*, G. Yeni-Komshian, J. Kavanagh, & C. A. Ferguson, eds., Academic Press, New York, pp. 217-228.
Jones, S. S. 1996, "Imitation or exploration? Young infants' matching of adult gestures", *Child Development*, vol. 67, pp. 1952-1969.
Lieberman, P. 1980, "On the development of vowel production in young children," in *Child Phonology: Vol 1, Production*, G. Yeni-Komshian, J. Kavanagh, & C. A. Ferguson, eds., Academic Press, NY, pp. 113-142.
Meltzoff, A. M. & Moore, M. K. 1997, "Explaining facial imitation: a theoretical model", *Early development and parenting*, vol. 6, pp. 179-192.
Messum, P. R. 2002, "Learning and teaching vowels", *Speak Out! (Whitstable, IATEFL)* no. 29, pp. 9-27.
Morrison, I. 2002, "Knowing our imitations", *Trends in Cognitive Sciences*, vol. 6, no. 3, pp. 115-116.
Smith, V. A., King, A. P., & West, M. J. 2000, "A role of her own: female cowbirds, *Molothrus ater*, influence the development and outcome of song learning", *Animal Behaviour*, vol. 60, pp. 599-609.
Underhill, A. 1994, *Sound Foundations* Heinemann ELT, Oxford.
Wolpert, D. M., Ghahramani, Z., & Flanagan, J. R. 2001, "Perspectives and problems in motor learning", *Trends in Cognitive Sciences*, vol. 5, no. 11, pp. 487-494.
Young, R. 1995, "Caleb Gattegno's "Silent Way": some of the reasons why", *Methoden der Fremdsprachenvermittlung (University of Mainz)* pp. 55-74.
Young, R. 2000, "Round table discussion (Nusbaum, Stevick, Thornbury, Young): Under what circumstances do we apply the word 'scientific' in language learning?", *Prism (Paris) - a learning journal*, vol. 5, pp. 7-183.