



HAL
open science

L2 pronunciation: Seven learning/teaching paradigms found in instructed learning

Piers Messum, Roslyn Young

► **To cite this version:**

Piers Messum, Roslyn Young. L2 pronunciation: Seven learning/teaching paradigms found in instructed learning. Alice Henderson; Anastazija Kirkova-Naskova. Proceedings of the 7th International Conference on English Pronunciation: Issues and Practices, pp.160-171, 2023, <10.5281/zenodo.8225633>. <hal-04178959>

HAL Id: hal-04178959

<https://hal.science/hal-04178959v1>

Submitted on 8 Aug 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons CC BY 4.0 - Attribution - International License

Messum, P & Young, R. (2023). L2 pronunciation: Seven learning/teaching paradigms found in instructed learning. In A. Henderson & A. Kirkova-Naskova (Eds.), *Proceedings of the 7th International Conference on English Pronunciation: Issues and Practices* (pp. 160–171). Université Grenoble-Alpes. <https://doi.org/10.5281/zenodo.8225633>

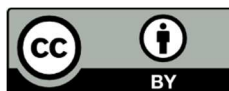
L2 pronunciation: Seven learning/teaching paradigms found in instructed learning

Piers Messum
Pronunciation Science Ltd.

Roslyn Young
Pronunciation Science Ltd.

For good production in L2 pronunciation, learners have to be able to form new sounds in many syllabic contexts and at speed, so pronouncing an L2 requires learning new motor skills. The basic unit of motor skill development is the action-perception cycle, within which there is a need for the learners' attempts at a target to be evaluated, by themselves or by an expert. We present a theoretical framework based upon how learning/teaching paradigms meet the need for such evaluation. This yields a taxonomy of seven paradigms.

Keywords: L2 pronunciation learning, taxonomy, L2 speech sound production, motor skill



This chapter is based on the workshop given by the authors at the 7th International Conference English Pronunciation: Issues and Practices (EPIP 7) held May 18–20, 2022 at Université Grenoble-Alpes, France. It is licensed under the Creative Commons Attribution 4.0 International License. To view a copy of the license, please go to: <http://creativecommons.org/licenses/by/4.0/>.

1 Introduction

For many years, we have been concerned that classroom-based pronunciation teaching has been developed, practised, and researched without sufficient prior investigation into how learners learn to pronounce an L2. Others have made similar points. For example, Foote and Trofimovich (2018) describe a lack of theory to guide L2 pronunciation research as being one of the most acute problems in the field. They explore how linguistic, psychological, interactionist, sociocultural, identity and sociocognitive perspectives can be useful for research, but acknowledge firstly that these have little to contribute to L2 pedagogy and secondly that this is a significant shortcoming, because, “research in L2 pronunciation should ultimately inform pronunciation teaching” (p. 85). For theory to support pronunciation teaching, it seems clear that it should address how learners learn to pronounce an L2.

This article starts to address one aspect of this issue: learning to produce L2 speech sounds which are not present in the L1 inventory (e.g., the production of English /l/ and /r/ by Japanese speakers, or French /y/ by English speakers). When we refer to learning new L2 speech sounds, we include learning new sounds in different contexts, not just in isolation (i.e., the learner producing sounds, sounds within clusters, sounds in words and then in phrases). Pronouncing an L2 sound, therefore, requires developing new motor skills.

It is obvious that one should try to understand how something is learnt before teaching it, so why is the mechanism used in learning to pronounce L2 so rarely discussed by researchers? Presumably the issue has seemed to be unproblematic: teachers and researchers know that some forms of vocal learning in speech and singing can be done by imitation (by which they mean a self-evaluated auditory matching-to-target process) and assume that this is possible for L2 speech sounds. They may think that children learn L1 pronunciation in this way, and that we can invoke the same mechanism in older learners, albeit with teaching practices that are adapted to take account of differences between the two age groups (summarised by Stevens, 1974).

The field in general, both researchers and teachers, has thus assumed that learners will develop L2 pronunciation if they are presented with exemplars to be copied. In so-called Intuitive-Imitative (I-I) approaches (Celce-Murcia et al., 2010; Kelly, 1969), this is imagined to be sufficient. In Analytic-Linguistic (A-L) approaches, this stage is preceded by training to improve how learners hear L2 and/or with phonetic information and rules which, it is believed, will become know-hows (automatised procedural knowledge) through repeated application.

We have identified six other ways in which it has been or could be imagined that L2 pronunciation is learnt/taught. To create this taxonomy, we approached the problem from two directions. First, we asked, “Since L2 pronunciation is a socially transmitted motor skill, what are the implications for how it is learnt?”, and then, “L2 pronunciation is taught in different ways; what are the learning mechanisms that can be inferred to underlie each of them?”

We start by outlining some conceptual points drawn from psychology and other disciplines which inform our approach (see §2). We describe the basis of our taxonomy (see §3), and then we describe the different learning mechanisms it identifies and the teaching practices which are based upon them (see §4–6).

2 Conceptual points for understanding pronunciation learning

Messum and Young (2021) gave reasons why learning to pronounce an L2 is “a highly unusual activity” (p. 170) among socially transmitted skills, including the fact that the actions involved are largely invisible and that production and perception skills developed for L1 pronunciation interfere with the process. Furthermore, we argue that five conceptual points

need to be taken into consideration if one is to think clearly about learning to pronounce an L2. These are presented one by one in the following sub-sections.

2.1 Pronunciation learning is not imitation

When teachers provide a spoken model for their learners in pronunciation classes, they are presenting them with the results of their actions and not the actions themselves, since most of the actions involved are hidden inside the mouth. In this way, teaching pronunciation is unlike teaching most skills, where a model shows the learner what the demonstrator is doing as well as the result. Metaphorically speaking, if a winner of the Masters were teaching golfers how to drive a ball by simply striking 300-yard shots off the tee, we do not think they would learn as much as if he gave them advice on how to improve their own swings. But at least they might pick up something from watching him in action. Now imagine if he were hitting those 300-yard drives while standing behind a tarpaulin so that the golfers could not see what he was doing and could only see the result: a ball sailing down the fairway every time he produced a 'model'. Then they would get very little, if anything, from the experience.

In technical discussions, *imitation* refers to copying of actions as well as reproducing the demonstrator's results. For this reason, the colloquial use of the word *imitation* for the result of a self-evaluated matching-to-target process in L2 pronunciation is inappropriate. The field should be using the word *emulation*, or more strictly, *goal emulation*: the adoption of the goals of the demonstrator, the reproduction of his/her results but not the copying of his/her actions (Call & Carpenter, 2002; Whiten & Ham, 1992). It is important for us to use terminology that acknowledges the complexity of the process of learning L2 speech sounds. In imitation, there are two sources of information that the learner can attend to at will: the actions of the model and the results obtained. In emulation, there is only one source, the results, and for L2 speech these are an acoustic signal whose interpretation by most learners is distorted by the mechanism of their L1 perception.

2.2 Two ways of listening: Autocentric and allocentric

There has been a longstanding scientific interest in the two products of our senses, sensation and perception, and in their relation. Humphrey (1992) opens his discussion of this by quoting Reid (1785): "The external senses have a double province – to make us feel, and to make us perceive" (p. 46).

In the auditory domain, Öhman (1975) gives the following example: we might experience a household event either as the sound of a refrigerator door shutting in another room or as meaningless "concrete music", that is not about the state of the world but is "an immediate awareness of the developing states of [our] auditory sense" (p. 42). Similarly, when listening to a speech sound, it is possible to attend to it in two different ways: as something meaningful or as noise. For instance, we might experience an event as a linguistic /p/ or as the meaningless noise of an explosion of air from the mouth. Since the first way of attending to sounds gives us potentially useful information about the outside world, it is the attentional set that we use almost all the time. Indeed, when meaning is available from a signal, it is often difficult to attend to it as noise.

In research on speech perception, various distinctions have been drawn between the attentional sets of listening, and a variety of terms have been used to describe them. There have even been proposals for three-factor models of speech perception, with auditory, phonetic, and phonemic components (Werker & Logan, 1985). However, to describe the particular distinction that Öhman (1975) drew and to avoid any possible confusion with other distinctions that have been made, we will use the terms proposed by Schachtel (1959), i.e.,

autocentric for awareness of sensations and *allocentric* for meaningful perception. These terms capture the point that the experience of mere noise is an awareness of the state of our own auditory sense (autocentric mode) while experiencing meaning is an awareness of the state of the outside world (allocentric mode).

It is necessary to make this distinction when considering pronunciation because speech sounds (instances of phonemes, or strings of phonemes) are noises to which we give linguistic significance in speech. When we learn a new L2 sound we are learning to produce a noise (or set of noises) that will be recognised as that particular sound by other speakers of the L2.

2.3 Two forms of memory: ASM and ALTM

Current models of memory distinguish short-term and long-term memory, and posit at least two components of short-term memory that deal with sound (Scott & Mishkin, 2016). The first is a phonological store that can be supplemented with subvocal rehearsal to sustain speech representations. However, this is not of any obvious use in learning to produce new L2 sounds.

The second is auditory sensory memory (ASM), which Nees (2016) describes as “a set of acoustic features organized in time that can be consulted to complete behavioural tasks, including comparing sounds to one another” (p. 1). There is a panoply of other names in use: the sensory register, echoic memory, acoustic short-term memory, passive short-term memory, pre-perceptual auditory memory, and so on. ASM is reported as decaying rapidly, usually within a few seconds (Nees, 2016).

Auditory long-term memory (ALTM) may also be relevant to learning new L2 sounds. As it is usually conceived, it does not store raw, sensory intake (noises). However, since selective attention is considered to isolate information in perception and then store it in long-term memory (Barsalou 1999), there does not seem to be a principled reason against any detail at any level being captured in ALTM (e.g., a novel feature of an L2 sound), provided the learner has noticed it.

2.4 Perceptual representations are not production representations

One determiner for what is captured in long-term memory is the current task, since a task is a driver for selective attention. If the task is to learn to identify or discriminate any type of object, then the abstraction will be optimised for that purpose, but not necessarily structured for recall or to guide subsequent production. Perceptual training in L2 pronunciation is usually evaluated by the success of learners in identifying or discriminating new L2 sounds. However, this cannot be taken to mean that they have developed a type of representation that can directly inform production.

To better appreciate this point, we suggest you stop reading and, from memory, draw the Ford Motor Company logo, trying in particular to get the style of the lettering right. Despite your undoubted ability to recognise this logo (which you have probably seen thousands of times), and to discriminate it from similarly coloured or shaped logos, you would be unusual if your production of the <f> or <r> matches the detail of the original. In any perceptual modality, to be able to recognise something does not mean one knows how to produce it.

2.5 Mirroring: Learning through others

Before video recording technology was available, springboard divers had to rely upon a coach to evaluate their performance and hence to improve. The coach acted as a form of

mirror to the learner, and psychology has demonstrated many ways in which we learn about ourselves through mirroring interactions. Moreover, it is very possible that children's L1 pronunciation develops this way, during vocal exchanges in which their caregivers imitate their babbling, reformulating it into well-formed L1 speech sounds (Messum & Howard, 2015). Munhall et al. (2021) lend support to this hypothesis, considering it to be consistent with current thinking on how speech perception develops in children.

Maas et al. (2008) use the terminology of Schmidt's (1975) schema theory (discussed below) to describe this mechanism:

Before the recognition schema can be used to judge the accuracy of the movement, the system must first learn which sensory consequences are to be considered "correct." There is often a clear reference of correctness (e.g., a golf ball must end up in the hole), but there are cases in which the reference of correctness is not directly available or interpretable to the learner but instead depends on feedback from an instructor, such as when learning to perform a somersault in diving. In such cases, the learner must calibrate the expected sensory consequences with an externally provided reference of correctness, so that the internal error signal may serve to correct errors on future trials without external feedback. (p. 279)

The reference of correctness for L2 speech sounds may be neither directly available to learners nor interpretable by them: *unavailable* because of the perceptual opacity of speech sounds (Heyes & Ray, 2000), with the learners hearing their own voice mixed in with bone-conducted sound and hearing what they expect themselves to be producing rather than the actual output (Munhall et al., 2021); and *uninterpretable* because of the effect of their L1 listening expertise on how they hear L2. In these circumstances, an externally provided reference of correctness becomes necessary.

3 A taxonomy of L2 pronunciation learning/teaching paradigms

Now that these conceptual points have been made, we present seven mechanisms for how the production of L2 speech sounds might be developed through conventional learning, with a note in most cases of an associated teaching practice or practices (Figure 1). A more comprehensive taxonomy will include learning mechanisms proposed in the literature that invoke innate, neural capacities (Messum & Young, in preparation).

Before discussing each mechanism in turn, we first explain why our main taxonomic question is how a speaker evaluates their attempts at L2 pronunciation.

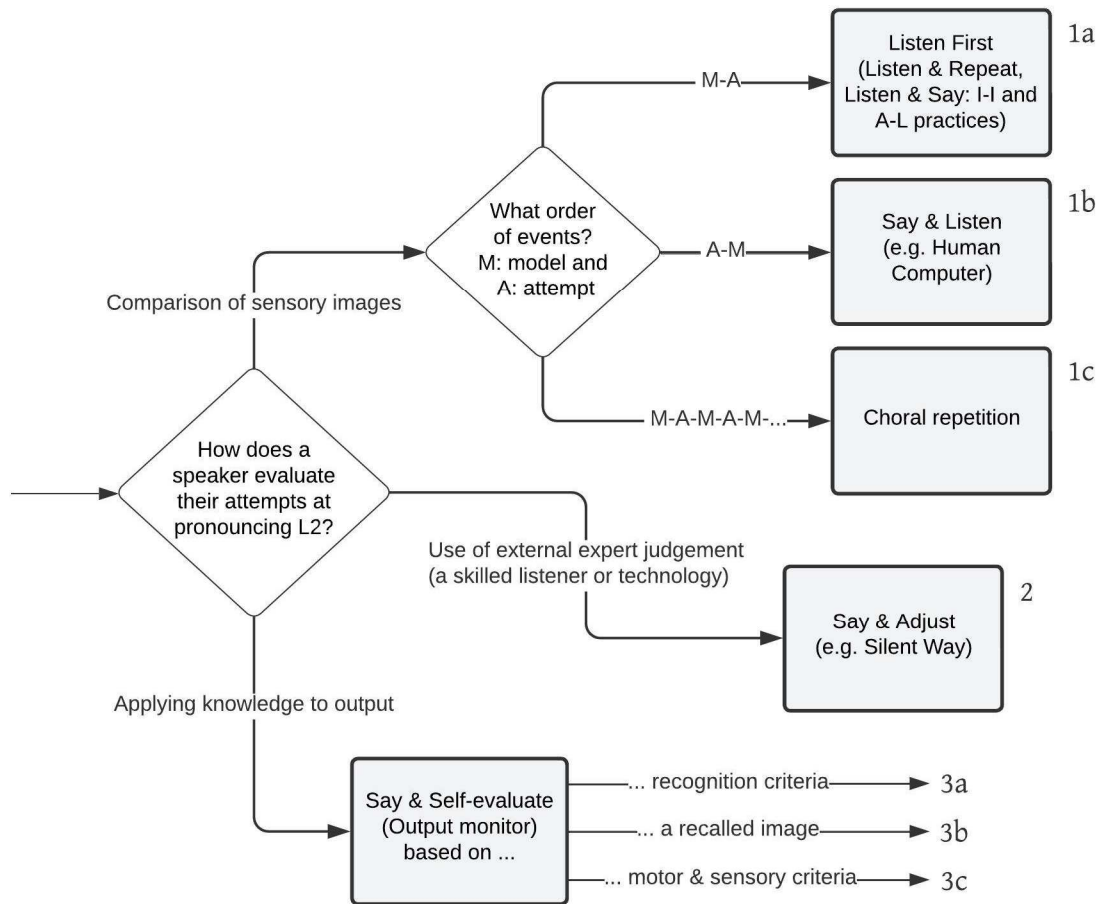
While the learning and teaching of L2 pronunciation can be considered from various perspectives, there should be a common aim: that learners pronounce L2 proficiently. Learners have to develop new willed actions — do something different with themselves (i.e., with their articulatory apparatus) — if they are to pronounce L2 differently from how they pronounce L1. Pronouncing is a motor skill. With this starting point, we can broaden our understanding of L2 pronunciation learning by reference to the established body of research on motor skill development.

The foundational theories in this field — Closed-loop Motor Learning (Adams, 1971, 1987) and Schema Theory (Schmidt, 1975; Schmidt & Lee, 2005) — agree that action and evaluation are two key elements in motor learning. These inform the memory trace and perceptual trace respectively in Adams' theory, and the recall schema and recognition schema in Schmidt's. In repeated action-perception (A-P) cycles (Cutsuridis et al., 2011), learners develop sensorimotor contingencies between their actions and the results these produce. The action part of the cycle relies on the learner having some control over what they do: the

ability to do the same thing again or to do something new. In the perception part of the cycle the learner evaluates the results of the action judged against the target. This evaluation can be performed either by the learner, or by a competent other person who communicates their evaluation.

Figure 1

A Taxonomy of Learning/Teaching Paradigms in Instructed L2 Pronunciation



Note. The labels within the numbered boxes describe what a learner does in each case, with the name of an associated teaching practice noted if appropriate. Abbreviations: I-I, Intuitive-Imitative; A-L, Analytic-Linguistic. M, Model; A, Attempt. (Thus ‘M-A’ means ‘model followed by learner attempt’.)

Before the first attempt is made, the learner has a (perhaps unspoken) question, “How can I do this?”. After the attempt, they need to answer the question, “How successful was that?”. If they continue, they then have to ask, “What am I going to do now?” and, after the next attempt, “Was that better or worse than last time?” and “What difference did the new thing I tried make?”.

To establish the possible learning mechanisms for L2 pronunciation, we focus on the perception/evaluation side of the A-P cycle. Thus, our taxonomic question on the left of

Figure 1 asks how the learner evaluates their attempts at pronouncing an L2. The possible ways in which this could be done are documented on the three branches which lead away from the central question and which describe the learner's mental activity:

1. The learner compares *sensory* images: the image they retrieve from their own attempt that they compare to a model they hear. One (or perhaps both) of these would be held in ASM.
2. The learner is informed about their attempt from another's evaluation of it — be it from a teacher, another speaker of the L2, or some form of technology.
3. The learner evaluates their attempt using *knowledge* previously acquired. This knowledge might be: a) auditory criteria for correctness for the L2 sound; b) a fine-grained exemplar of the sound in ALTM that they use for comparison; or c) motor and sensory (proprioceptive) criteria for correctness. The knowledge needed for a) and b) would have been acquired through listening; that for c) would have been acquired from previous cycles of A-P learning.

Note that any actual classroom interaction might give the learner more than one source of information about their performance, but the sources will be drawn from these three basic categories.

The final termination points for the three branches from the central question in Figure 1 are seven learning/teaching paradigms for L2 pronunciation, labelled 1a to 3c. The boxes describe what the learner does in each case, with the name of an associated teaching practice noted if appropriate. The next sections (§4–6) explain the paradigms resulting from each of the three branches.

4 Paradigms 1a, 1b, and 1c: Comparison of sensory images

4.1 Listen First: Intuitive-Imitative and Analytic-Linguistic approaches (Paradigm 1a)

If the learner evaluates their performance based on a comparison of sensory images (the aural image created by their attempt at a sound and an aural image they hear as a model) then the most familiar classroom order of such events (and the usual order of such events in language learning apps on mobile phones) is that of Listen & Repeat: a model followed by a learner attempt. This is the exercise that best characterises Intuitive-Imitative (I-I) pronunciation teaching.

However, learners complain that they cannot *hear* the model, i.e., that they are not aware that the model L2 token is different from some similar sound in L1. Nothing can be achieved by asking them to copy what they do hear in these circumstances, so teachers seek ways to remedy this. Hence Analytic-Linguistic (A-L) approaches supplement the supposed natural mechanism of imitation with phonetic information and/or preparatory perceptual training. A-L practices are considered to complement I-I practices rather than replace them (Celce-Murcia et al., 2010). However, to acknowledge that the learner's response after A-L preparation is likely to be more carefully considered than what they can attempt in simple Listen & Repeat, we call it Listen & Say.

Whether or not there is a preparatory phase, notice that the principal learning move for production in both I-I and A-L approaches is for the learners to match an L2 sound which has been presented to them. These two Listen First approaches aim at a learner's ASM, through autocentric mode listening. Most current classroom practice is therefore based within this single paradigm in our taxonomy.

4.2 Say & Listen (Paradigm 1b)

It is possible to reverse the order of events in Listen First approaches, i.e., for the model to follow the attempt. For example, in the Human Computer technique from Community Language Learning (Curran, 1976), the teacher repeats the learner's sound, word or phrase in their own, correct L2. The learner is then not allowed to repeat the phrase after the teacher.

This prohibition on repetition prevents reflexive copying where the learner is not aware of their speech motor activity, i.e., what they themselves are doing to pronounce. Instead, there is a period of silence, in which the learner can compare the image they have of their own production to what they have just heard. This gives them the opportunity to evaluate the former and consider what they might change about it. When enough time has passed for the teacher's utterance to have faded from the learner's ASM, the learner is allowed to say the phrase again, informed by what they learned from the previous cycle. The teacher will repeat the phrase again in good quality L2, and the learner has the opportunity to compare and contrast the two renditions.

4.3 Choral repetition (Paradigm 1c)

In choral repetition, the teacher's model comes neither before nor after the learner's attempt, but rather in both places. In repeated, rapid sequences of model – attempt – model – attempt, it becomes unclear which of the mechanisms in boxes 1a and 1b describes the learning episode. Choral repetition of this type could involve either of them, depending upon how the learner chooses to direct their attention, potentially switching between the two more than once during any session.

Kjellin (2004) describes his rationale for choral practice led by the teacher, and his own classroom methodology that implements this. Based upon extensive classroom experience, he insists that 10, 20 or 30 repetitions of a phrase are inadequate and counterproductive, but that a block of 50-100 repetitions leads to success. Jones (2018) describes her own rationale and gives further advice on the use of choral repetition.

Immonem et al. (2022) report success training 7-year-old children to produce new L2 vowels using this learning/teaching paradigm: automatic, alternating presentation of two words containing the vowels with a fixed interstimulus interval of three seconds, over four training sessions of 30 pairs of repetitions. With similar training regimes, older children and adults have also been successful, but more slowly. It might be that younger children are better at attending to sounds as sensory objects and to be less deeply in the *grip* of L1 (Underhill, 2013).

5 Paradigm 2: Use of external expert judgement

In Paradigm 2, Say & Adjust, learners rely on the evaluation of their performance by an expert other: a metaphorical *mirror*. Within a mirror learning mechanism, it is acknowledged (and unproblematic) that the learners do not yet know the target: either how to create it, what it feels like or even what it sounds like. They only have some approximate idea that enables them to start production. With feedback from an expert, they can modify what they do, all the while getting to know the target better.

Sixty years of classroom practice has demonstrated the efficacy of such a process in L2 pronunciation teaching, starting with Gattegno's first version of the Silent Way and developed by him and other teachers since (Messum & Young, 2021). The paradigm need not be associated with the Silent Way. The empirical evidence presented by Warsi (2001) suggests that he was considerably more successful than Bradlow et al. (1997) in teaching the

production of English /l/ and /r/ to Japanese learners through Say & Adjust. Furthermore, the source of evaluation need not be human; technological evaluation and feedback can also be used (e.g., Kartushina et al., 2015; Sakai, 2016).

While the Say & Adjust paradigm is consciously deployed in Silent Way classes, it may also appear in conventional classes which seemingly employ a Listen First approach. When a teacher says, “Make a [...]” and the learner takes what they hear to be a *cue* rather than a *model* (and this distinction is key), they ignore the possibility of copying what the teacher has said and produce their current best attempt at the sound. Then, when the teacher gives feedback on how well the learner has done, informing a new attempt, this is a Say & Adjust exchange.

6 Paradigms 3a, 3b, and 3c: Applying knowledge to output

Three other paradigms involve self-evaluation based upon knowledge acquired previously. This notion can also be found in the field of Child Phonology. For example, based on developmental data, Kuhl (2000) described one way in which young children may learn to produce L1 speech sounds: “... early in life, perceptual representations of speech are stored in memory. Subsequently, these representations guide the development of motor speech” (p. 11854).

An analogue of this for L2 was described by Bradlow et al. (1997): “... perceptual learning leads to more accurate internal acoustic representations of the target speech sounds, and these improved representations function as acoustic templates that play an important role in monitoring the articulatory output” (p. 2307). They called such templates *output monitors*.

A monitor of some type, operating to evaluate the speaker’s own output, could rely upon any of three distinct types of previously acquired knowledge: recognition criteria, a recalled image, or motor and sensory criteria.

6.1 Self-evaluation based on recognition criteria (Paradigm 3a)

The speaker might have developed criteria to identify and/or discriminate L2 sounds produced by others (in isolation or in words) which they now apply to evaluate their own output. To return to the visual analogue involving the Ford logo, a novice illustrator trying to reproduce it without a model to copy might evaluate their attempt by wondering whether it looks right.

6.2 Self-evaluation based on a recalled image (Paradigm 3b)

The speaker might have developed an image of the L2 sound which is now stored within ALTM. They could compare their output against a version of this that they recall. With respect to the Ford logo, the novice in this case would be able to evoke (or recall) a detailed mental image they had created to guide their attempt.

6.3 Self-evaluation based on motor and sensory criteria (Paradigm 3c)

During previous work on the L2 sound, the speaker might have developed some or all of three different types of criteria that now help them during independent practice. They will know more about: 1) how (gesturally) to make the sound (i.e., what actions they need to perform); 2) how it feels to make the sound; and 3) how it sounds to them to make the sound.

These criteria alone do not allow them to exceed their best previous level of production accuracy, but they do allow them to consolidate their learning: producing the sound in different contexts, at different rates, with greater automaticity, etc.

Returning to the Ford logo, in this case the novice would have already drawn the logo several times, checking their attempts against an original, and hence developing facility and accuracy in its production. Now they are making a further attempt.

7 Conclusion

Our taxonomy is based on the characteristics of the perception side of the A-P cycle; specifically, the mechanism by which a learner evaluates their successive attempts. We have identified a number of possible evaluation mechanisms and the teaching approaches that rely upon them which, considered together, we call learning/teaching paradigms. This taxonomy will be expanded (Messum & Young, in preparation) to include paradigms that do not rely upon instruction, and we will also discuss the likely effectiveness of each of them.

As suggested by Pennington (2021): "... approaches that challenge standard practices on theoretical grounds deserve to be systematically investigated and their effects and effectiveness compared to those of the standard practices" (p. 17). The taxonomy presented here shows that there are plausible learning/teaching paradigms other than Listen First and technology-driven approaches. We support Pennington's call for researchers to investigate the effectiveness of these currently non-standard practices.

Acknowledgements

We would like to thank our two anonymous reviewers from EPIP7, as well as Marc-Georges Nowicki, Alice Henderson, Dan Frost, Anastazija Kirkova-Naskova, Martha Pennington, Jonathan Marks, and Adrian Underhill.

References

- Adams, J. A. (1971). A closed-loop theory of motor learning. *Journal of Motor Behavior*, 3(2), 111–150. <https://doi.org/10.1080/00222895.1971.10734898>
- Adams, J. A. (1987). Historical review and appraisal of research on the learning, retention and transfer of human motor skills. *Psychological Bulletin*, 101(1), 41–74. <https://doi.org/10.1037/0033-2909.101.1.41>
- Barsalou, L. W. (1999). Perceptual symbol systems. *Behavioral and Brain Sciences*, 22(4), 577–660. <https://doi.org/10.1017/S0140525X99002149>
- Bradlow, A. R., Pisoni, D., Akahane-Yamada, R., & Tohkura, Y. (1997). Training Japanese listeners to identify English /r/ and /l/: IV. Some effects of perceptual learning on speech production. *Journal of the Acoustical Society of America*, 101(4), 2299–2310. <https://doi.org/10.1121/1.418276>
- Call, J., & Carpenter, M. (2002). Three sources of information in social learning. In K. Dautenhahn & C. L. Nehaniv (Eds.), *Imitation in animals and artifacts* (pp. 211–228). MIT Press.
- Celce-Murcia, M., Brinton, D., & Goodwin, J. M. (2010). *Teaching pronunciation: A course book and reference guide* (2nd ed.). Cambridge University Press.
- Curran, C. (1976). *Counseling-learning in second languages*. Apple River Press.
- Cutsuridis, V., Hussain, A., & Taylor, J. G. (Eds.). (2011). *Perception-action cycle: Models, architectures and hardware*. Springer.

- Foote, J. A., & Trofimovich, P. (2018). Second language pronunciation learning. In O. Kang, R. I. Thomson, & J. M. Murphy (Eds.), *The Routledge handbook of contemporary English pronunciation*. Routledge.
- Heyes, C. M., & Ray, E. D. (2000). What is the significance of imitation in animals? *Advances in the Study of Behavior*, 29, 215–245. [https://doi.org/10.1016/S0065-3454\(08\)60106-0](https://doi.org/10.1016/S0065-3454(08)60106-0)
- Humphrey, N. (1992). *A History of the mind: Evolution and the birth of consciousness*. Simon & Schuster.
- Immonen, K., Alku, P., & Peltola, M. S. (2022). Phonetic listen-and-repeat training alters 6–7-year-old children’s non-native vowel contrast production after one training session. *Journal of Second Language Pronunciation*, 8(1), 95–115. <https://doi.org/10.1075/jslp.21005.imm>
- Jones, T. (2018). Repeat after me. *As We Speak (TESOL SPLIS Newsletter)*, March. <http://newsmanager.commpartners.com/tesolsplis/print/2018-03-15/3.html>
- Kartushina, N., Hervais-Adelman, A., Frauenfelder, U. H., & Golestani, N. (2015). The effect of phonetic production training with visual feedback on the perception and production of foreign speech sounds. *The Journal of the Acoustical Society of America*, 138(2), 817–832. <https://doi.org/10.1121/1.4926561>
- Kelly, L. G. (1969). *25 Centuries of language teaching*. Newbury House.
- Kjellin, O. (2004). *Choral Practice – The Neurophysiological Opportunist’s Way*. https://www.academia.edu/2184625/Choral_Practice-the_Neurophysiological_Opportunists_Way
- Kuhl, P. K. (2000). A new view of language acquisition. *Proceedings of the National Academy of Sciences USA*, 97(22), 11850–11857. <https://doi.org/10.1073/pnas.97.22.11850>
- Maas, E., Robin, D. A., Austermann Hula, S. N., Freedman, S. E., Wulf, G., Ballard, K. J., & Schmidt, R. A. (2008). Principles of motor learning in treatment of motor speech disorders. *American Journal of Speech-Language Pathology*, 17(3), 277–298. [https://doi.org/10.1044/1058-0360\(2008/025\)](https://doi.org/10.1044/1058-0360(2008/025))
- Messum, P. R., & Howard, I. S. (2015). Creating the cognitive form of phonological units: The speech sound correspondence problem in infancy could be solved by mirrored vocal interactions rather than by imitation. *Journal of Phonetics*, 53, 125–140. <https://doi.org/10.1016/j.wocn.2015.08.005>
- Messum, P., & Young, R. (in preparation). L2 pronunciation: a taxonomy and assessment of nine learning/teaching paradigms.
- Messum, P., & Young, R. (2021). Teaching students to pronounce English: A motor skill approach in the classroom. *RELC Journal*, 52(1), 169–178. <https://doi.org/10.1177/0033688220964107>
- Munhall, K. G., Cui, A., O’Donoghue, E., Lamontagne, S., & Lutes, D. (2021). Perceptual control of speech. In J. S. Pardo, L. C. Nygaard, R. E. Remez, & D. B. Pisoni (Eds.), *The handbook of speech perception* (pp. 97–121). Wiley. <https://doi.org/10.1002/9781119184096.ch4>
- Nees, M. A. (2016). Have we forgotten auditory sensory memory? Retention intervals in studies of nonverbal auditory working memory. *Frontiers in Psychology*, 7. <https://doi.org/10.3389/fpsyg.2016.01892>
- Öhman, S. E. G. (1975). What is it that we perceive when we perceive speech? In A. Cohen & S. Nooteboom (Eds.), *Structure and process in speech perception* (pp. 36–47). Springer.
- Pennington, M. C. (2021). Teaching pronunciation: The state of the art 2021. *RELC Journal*, 52(1), 3–21. <https://doi.org/10.1177/00336882211002283>
- Reid, T. (1785). *Essays on the intellectual powers of man*. Longman.
- Sakai, M. (2016). (Dis)connecting perception and production: Training adult native speakers of Spanish on the English /i/–/ɪ/ distinction [Unpublished doctoral dissertation]. Georgetown University.
- Schachtel, E. (1959). *Metamorphosis: On the development of affect, perception, attention, and memory*. Basic Books.
- Schmidt, R. A. (1975). A schema theory of discrete motor skill learning. *Psychological Review*, 82(4), 225–260. <https://doi.org/10.1037/h0076770>
- Schmidt, R. A., & Lee, T. D. (2005). *Motor control and learning: A behavioral emphasis* (4th ed.). Human Kinetics.
- Scott, B. H., & Mishkin, M. (2016). Auditory short-term memory in the primate auditory cortex. *Brain Research*, 1640, 264–277. <https://doi.org/10.1016/j.brainres.2015.10.048>

Messum & Young
Seven learning/teaching paradigms

- Stevens, P. (1974). A Rationale for Teaching pronunciation: The rival virtues of innocence and sophistication. *ELT Journal*, 28(3), 182–189. <https://doi.org/10.1093/elt/XXVIII.3.182>
- Underhill, A. (2013). Cinderella, integration and the pronunciation turn. *Speak Out!*, 49, 4–8.
- Warsi, J. (2001). Effects of visual instruction on second language productive phonology [Unpublished doctoral dissertation]. Boston University.
- Werker, J. F., & Logan, J. S. (1985). Cross-language evidence for three factors in speech perception. *Perception & Psychophysics*, 37(1), 35–44. <https://doi.org/10.3758/BF03207136>
- Whiten, A., & Ham, R. (1992). On the nature and evolution of imitation in the animal kingdom: Reappraisal of a century of research. In P. J. B. Slater, C. Beer, M. Milinski, & J. S. Rosenblatt (Eds.), *Advances in the study of behavior*, Vol. 21 (pp. 239–283). Academic Press.

About the authors

Piers Messum has taught English in Japan, France, and the UK. His PhD in Phonetics from University College London investigated how children learn L1 pronunciation. He co-authored *How We Learn and How We Should be Taught* (2011) with Roslyn Young. He now teaches in London, and does teacher training and materials development for Pronunciation Science Ltd.

Email: p.messum@gmail.com

Roslyn Young worked at the University of Franche-Comté in the Applied Linguistics Centre from 1968 until she retired, teaching spoken English and sometimes spoken French in intensive courses. She ran her first ‘pronunciation clinic’ in 1978 and has worked continuously since then on ways of improving the teaching of pronunciation. She wrote her doctoral thesis on Caleb Gattegno’s model of learning and its relevance to his work in language teaching. She has published numerous articles on the teaching of pronunciation and on how people learn.

Email: roslynyoung@gmail.com