Five Cornerstones for Second-Language Acquisition - the Neurophysiological Opportunist's Way

Olle Kjellin¹
M.D., Ph.D.

Introduction and Summary:
Recent research into language acquisition and into the neurosciences have revealed interesting, fundamental similarities between first and second language acquisition, even on a primary, neurobiological level. Based on some of these similarities, an attempt is offered here at an interdisciplinary synthesis of linguistic and neurobiological research results. This paper will advocate a focus, in second-language (L2) teaching and learning, on five factors that are essential cornerstones for the normally perfect success of first-language (L1) acquisition, viz. hearing, prosody, statistics, categorical perception, and compensatory articulation. Suggestions are offered how these factors can be pedagogically adapted to the adult learner's situation and implemented in the classroom as strategies for a higher-than-usual ultimate attainment of L2 competence and performance, particularly pronunciation. This interdisciplinary synthesis is based on the author's training as a speech scientist and 30+ years of experience as a second language teacher (with a particular interest in pronunciation), interwoven with medical training and practice in radiology (with a focus on the speech and swallowing apparatus), in audiology, and in the neurosciences involving learning and its opposite, the dementias.

Cornerstone #1: Hearing
The auditory system is the only channel for the acquisition of speech and a spoken language. Hearing also is one of the first senses to mature, to be "ready" to use. The process commences in utero some three months before birth, when the cochlea has actually reached its adult size, and fetal hearing slowly begins functioning. Newborn infants have "continuous" perception. This means, as numerous ingenious experiments have shown, that they can auditorily distinguish all the world's speech sounds so far tested, such as the small difference between a French and an English /p/. This is because their language acquisition begins from scratch with no preconceived opinion as to what sound qualities are to be expected. This is in contrast to older children and adults, who are "locked up" by the specific sound categories that occur in their ambient language(s). (See Cornerstone #3, below.) This auditory, phonological petrifaction into the linguistic/phonological categories of the first language begins within weeks after birth and is more or less complete by the end of the first year of life, making us practically "deaf" to the sound categories of other languages from that time on.

However, an important consolation is that this deaf-like state, called amblyacusis, is no real deafness but only a diminished sensitivity that can be revitalized by specific training. It has not been shown that there exist any "critical age" around puberty (Birdsong, 1999), as is often claimed. Rather, there are several different sensitive periods for the various components of speech and language. The sensitive period for the L1 phonological development is about twelve postnatal months (not twelve years!). The sensitive period for the development of semantic hierarchies and relations is up to about 4-5 years of age, and for syntax up to the late teens (Ruben, 1997).

These figures, however, have been found for L1 acquisition. No limits have been found for further learnability later on in life. Or it would be illogical to postpone, for instance, car driving to after age 18, or skilled professional training such as neurosurgery to age 25 or even later. The end result mainly depends on the quality and quantity of training, and surprisingly little of the individual differences in ultimate attainment depends on a genetic contribution (Ericsson, Krampe, & Tesch-Römer, 1993). Indeed, it has not been shown at all that spoken

¹ Author's affiliation: Memory Clinics, Dept. of Geropsychiatry, Adult Psychiatric Hospital, Vaxjo, Sweden; and (part-time) Dept. of Swedish as a Second language, Faculty of Humanities, Vaxjo University, Vaxjo, Sweden. E-mail: olle.kjellin@home.se.
language learning substantially differs from any other motor and intellectual learning. Normal prelinguistic toddlers spend their entire waking time practicing communicative skills of all sorts in an optimal learning environment. The result can only be "perfect". Adult elite learners, too, spend an average of about four hours a day on deliberate practice of whatever is their profession or sport, e.g. painting, music, typewriting, chess, tennis, pole-vaulting, etc. (Ericsson et al., 1993). The need for extensive practice is commonsense in all fields requiring any kind of advanced skills (Hallam, 1997). Except in language education, where it went out of fashion decades ago with the death of the Audio-Lingual Method (ALM) (Littlewood, 1999).

Normal L1 speakers keep "practicing" their mother tongue hours and hours on end by sheer every-day usage. No wonder we all become "elite" speakers. How much do average L2 learners practice? (Bongaerts, 1999)

Strategy #1: Extensive Listening and Hearing Exercises
Commence a new language course with extensive listening and hearing exercises. To overcome the language-specific amblyacusis, the teacher points out the new, relevant details, the new categories, and, above all, the boundaries between them. Make the learners aware of the salient details that "make it sound English", for instance the particulars of the length of stressed syllables. Physically and physiologically, there are no differences at all between the ways various speech sounds of any language make an impact on the eardrums, ossicular chains and cochlear structures. But the brain has to be cognitively trained to detect, identify, and interpret the new details.

A quite recent, very interesting and pedagogically useful finding about the brain is that there are direct neural connections from the auditory centres in the temporal lobes to motor areas in the frontal lobes, motor areas responsible for facial and oral movements (Kohler et al., 2002; Rizzolatti, Fadiga, Gallese, & Fogassi, 1996). These motor neurons are active in speech, of course, but in addition they are also active when we hear speech - at least if it is such speech that we can understand and that is relevant to us. Actually, it seems as if normal speech perception is more or less dependent on a pattern recognition mechanism, by which the incoming speech signal is mirrored in the motor neurons that would have been active in the production of the same sounds and stretches of speech. Such "mirror neurons", also called "imitation neurons", were first found in other primates, and their degree of motor activity was directly associated with the relevance of the actions mirrored (such as seeing the experimenter's lifting a peanut - high activity - or just an empty hand - no activity) (Ferrari, Gallese, Rizzolatti, & Fogassi, 2003). These new insights lend strong support to the Motor Theory of Speech Perception (Liberman & Mattingly, 1985; Liberman & Whalen, 2000).

The effect of incoming vocal sounds on the listener's own vocal apparatus is so strong, that if you look at a person who is listening very attentively to what you are saying, you will see slight lip movements matching your own lip movements, though with a little time delay (that reveals the time taken to "perceive" your message). Also, everybody surely has had the experience of listening to a hoarse speaker and feeling an urge to clear your own throat!

This strong effect of incoming vocal sounds on the listener's own vocal apparatus can be exploited in the classroom: Choral practice. (More practical details are given in [Kjellin, this issue]). In choral practice dominated by the teacher, everybody in the group of learners will be more or less "forced" to get the same rhythm and melody as the teacher, automatically and seemingly magically and to everybody's great surprise and satisfaction. Actually, this is no more magic than the fact that it is even difficult for most people in a choir to sing a different tune than the rest of the choir. The same phenomenon, of course, arises in a speaking chorus too.

First, the teacher should pronounce the target phrase 7 times alone to give the learners the chance to grab the whole phrase in their working memory. By the 3rd or 4th time, the teacher may see that many lips are already moving along, silently. By the 7th time, every mouth is ready to try. So next, the group speaks in chorus with the teacher some 10-15 times. Then the teacher speaks alone again some 5-10 times, allowing for new "discoveries" by the learners and prompting them to come back into chorus when they feel mature for it. This procedure is repeated ad lib. Later on individual practice ensues, but should also alternate frequently with choral practice. The choral practice efficiently and effectively co-trains both perception and production and establishes audio-motor links between the two, a kind of "inner ear" in long-term memory that will provide templates to guide and filter future independent speech production.

During the exercises, one (only one) important detail of the prosody or pronunciation that "makes it sound English" should be explicitly shown and exercised until it "settles". Then the next detail until it settles, etc. (It is
important not to overload the learners with details, they will have plenty to concentrate on anyway.) The details to be highlighted should typically involve one of three areas: (i) stressed syllable length for the rhythm, (ii) unstressed syllable reductions (including vowel neutralization, final-consonant death, and "liaison", the syllable-final consonant transmigration to become syllable-initial whenever the next syllable lacks an initial consonant), also for the rhythm, and (iii) quite often the shape of the lips, rounded or spread. These crucial three areas are mnemotechnically collected as "the 3 R:s", Rhythm, Reductions, Rounding. The teacher will soon notice that a working focus on the 3 R:s will improve the learners' pronunciation more than anything else.

Cornerstone #2: Prosody

Prosody, the rhythm and melody of speech, has been shown to be pivotal to normal language acquisition and use. It functions as the "skeleton", or framework, on which to build not only pronunciation, but also vocabulary, morphology, word order, discourse rules, etc. (Gerken & McGregor, 1998; Hung & Peters, 1997). Prosody thus also is the "lubricating oil" for normal, spoken communication. It will be the very fundamental instrument, or even medium, for our spoken communicative skills. Neuropsychologically, the rhythm of speech aids the receiving brain in its speech anticipation processes. The brain is a "prediction machine", and much of what we hear spoken to us will be locally predictable from the pragmatic, presuppositional, semantic, syntactic, and morphological context of the moment. The present reader may verify this by recalling how you usually can fill in the correct forms of the correct words whenever your interlocutor loses the thread and suddenly becomes silent during a conversation. Acoustically, large parts of the various contexts are built into the rhythm of speech. Indeed, interlocutors with normal social competence even tend to synchronize to each other's speaking rhythm, seemingly as an intuitive measure to facilitate turn-taking (Cowley, 1997). If turn-taking is desired. A sudden change of tempo may silence the other party.

Perception and comprehension involve (at least) three steps: detection, identification, and interpretation. The cerebral cortical areas responsible for these processes have their language-specific Eigenrhythmus. This is an intrinsic, spontaneous stand-by rhythm, ready to catch and lock on to a speaker's rhythm that we happen to hear (Greenberg, 1995), thereby enabling the immediate and even anticipatory speech perception that is characteristic of ordinary conversation, including the famous "cocktail party effect", in which one speaker's voice can be robustly selected in the midst of a cacophony of many other voices (Teder, Kujala, & Näätänen, 1993).

If, however, the speaker's prosody differs too much from what the listener expects, the lock-on mechanism may malfunction and the perception processes will take longer time, measurable in tens or hundreds of milliseconds per phrase. This is because the mirror neurons are unaccustomed to this prosody, and it will put a larger workload on the listener's working memory as well as on associative neural networks. It may, as a result, even make him or her feel a bit tired after a longer conversation with somebody having a very different prosody. This is a purely neurophysiological reality and has nothing to do with the listener's emotions or attitudes towards the speaker. This is not to deny that some listeners' emotions and attitudes may be secondarily affected by such experiences, usually negatively. And that is one important rationale for the present author's focus on prosody with the hope of reducing xenophobia by making immigrant speech more communicative on these basic and primary levels of neurophysiology and neuropsychology.

All human prosodies are fundamentally the same - a rhythm and a tone contour. Prosody is produced by the voice organ (the larynx with its vocal folds) and the lungs with their driving force, the air stream. These structures are identical in all nationalities and already fully functional at birth, primarily for cries and airway protection. Thanks to this inborn functionality, the infant will acquire its ambient prosody (or prosodies, if in a multilingual environment) from very early on. Its first word by around age 12-18 months is never, ever pronounced with any kind of "foreign accent"! By the time the toddler knows 25 words its prosody is mature and adultlike (Vihman & DePaolis, 1998). Thus, although it takes some 4-6 years before the child can produce clear speech (due to anatomical and neurological articulatory constraints affecting vowels and consonants), no normally developing child will ever be regarded as having a "foreign accent" in its first language(s). A "foreign accent" is mainly a prosodic problem.

However, some particulars of prosody are specific to each language, each dialect, and each sociolect, such that the "wrong" prosody, e.g., a distant dialect or a foreign accent, may sometimes become a spoke in the wheel, like sand in the machinery rather than the expected "lubricating oil". This is usually not an insuperable obstacle, but an unnecessary distraction making communication less smooth, tiring the listener, and sometimes causing irritation, and at times even jeopardizing mutual understanding.
Initially, the fetus in the womb is only capable of hearing rather low frequencies, or tones. But this is exactly what is needed to hear the prosody. The mother's voice is well transmitted directly from the voice organ through the body tissues to the fetus. Vowels and consonants, on the other hand, are only formed when the air passes out of the speaker's mouth or nose. Airborne sounds from the outside world to the womb are greatly attenuated by the air-tissue interface at the abdominal wall. Moreover, vowels and consonants contain many higher frequency components that are less well picked up by the fetus anyway.

**Strategy #2: Primarily Practice Prosody**

Make prosody practice a prioritized primary target for both listening and speaking exercises. Infants get it by lack of choice, i.e., Mother Nature's Wisdom. Adults in beginner's courses (and their teachers) have to choose deliberately to focus on things prosodic and avoid confusing the learning situation with as yet rather irrelevant talk about vowels and consonants, word forms, spellings, meanings, synonyms, etc., etc. (Derwing & Rossiter, 2003) (Cf. the child's L1 situation.) Instead, humming, using a kazoo, and standing up, move around, and include gestures and body language are good ideas. In a language such as English, where syllable length is highly relevant for the rhythm and stress patterns, reinforcing it visually with the help of a rubber band that is stretched in synchrony with the long (stressed) syllables will significantly increase the interhemispheric communication between the linguistic, analytical left hemisphere and the holistic, pattern-oriented right hemisphere, thereby multiplying the neural networks built up by these exercises and optimizing the end result (Gilbert, 1999).

**Cornerstone #3: Statistics**

The central nervous system develops in response to incoming stimuli, all incoming stimuli, and nothing but the incoming stimuli. Extensive neural networks build up and are reinforced by each repetition of a particular stimulus. "Number of reoccurrences" is statistics. On an average, an adult neuron receives input from about 10,000 other neurons, and in its turn outputs to 10,000 other neurons, including feedback loops to itself and to many of its input neurons. There are some 100 thousand million neurons in an adult brain. Thus, the whole world's Internet, having less than 10 thousand million connected computers, is only a blind alley compared with the fantastic networks of just one single human brain.

Connections between nerve cells are called synapses. An incoming stimulus causes synaptic "budding" on twigs on the affected neurons, buds that will degenerate and disappear if the stimulus is not further repeated. But after scores of repetitions, fully functional synapses will develop in a vast network - a long-term memory trace has been established. On disuse, even this may eventually degenerate and disappear, but with continuous use it will remain for a life-time and even develop further and be a support for other memories and neural networks to build up.

Actually, a new-born brain has about twice the number of neurons than the adult, but quite few synapses per neuron. The nervous system develops by weeding out unused neurons and branching the utilized ones (arborization) to make room for an ever increasing number of synapses and neural networks.

So, in effect, the neural mechanism of learning is very similar to walking on a lawn: Tracks will form wherever you walk sufficiently many times. But nowhere else. And no tracks will form independently of other tracks. Rather, already existing tracks are a requirement for new ones to form. Every new piece of learning is, and must be, a variation of previous knowledge or skill.

Unused tracks may eventually be grown over. Long-term memory and motor skills are directly related to the quantity and quality of repetitions. Statistics. Deliberate practice is, literally, the mother of learning (Ericsson et al., 1993). Particularly, motor skills have interesting properties: Automatized skills are stored as "procedural memories", which are more or less ineradicable, even in advanced Alzheimer's disease (Ojemann, Buckner, Corbetta, & Raichle, 1997).

Furthermore, in language acquisition, whether L1 or L2, there is an important word-statistical component. The 15 most frequent words in a language will account for about 25 percent of all the words in an ordinary text. Similarly, the first 50 words will account for 40 percent, the first 100 words for 60 percent, the first 1,000 words for 85 percent, and the first 4,000 words for 97 percent. These figures hold for any language (Crystal, 1995), though some minor variations have to be expected. What is most important and interesting about this is the fact
that all or almost all of the 50 most frequent words are grammatical function words, such as the, of, to, in, and, a, for, was, is, ..., etc. This means that the L1 child can get his grammar right by sheer statistical reasons (Saffran, Aslin, & Newport, 1996). An adult brain will be at least as apt as the infant brain, presumably much more, thanks to its greater degree of maturity.

**Strategy #3: Practice Plenty of Times**

The simple motto here would be, "Most Heard, Best Remembered". If one reads an ordinary novel of, say, 70,000 words, the 50 most common words will be seen in 40 percent of those, i.e., 28,000 times, or 560 times each on an average, with the proper inflectional and syntactical context and all. It will be difficult to forget them! Thus, extensive reading of books and listening to the radio and ambient speech, without any ambition to understand everything but with the ambition to recognize as much as possible, will serve as the non-child L2 learner's substitute for the naturally superstatistical environment of the L1 learner.

For the automatization of speaking and pragmatic skills, extensive repetitions of listening and pronouncing exercises of whatever phrases that happen to be at hand will make a magical trick. Whole phrases are preferred over single words for several reasons: A phrase with at least three stressed syllables will make up a good rhythm, and also contain the pertinent syntactical connections. Some 50 to 100 repetitions per practice phrase will be the minimum, particularly for beginners during the first 3-4 weeks, while this very "instrument", L2-English speech, is learnt. Less than that will be a waste of time, because the neural "tracks" will be too faint and soon lost. With modern brain imaging techniques it has been shown that it takes "up to 15 minutes" to change the cerebral activity from a "newbie" pattern to an "automatized" pattern (Raichle et al., 1994).

In the present author's experience, the limiting factor is the teacher's patience, not the learner's zeal, inclination or disposition. The teacher has to be aware of the fact that although it may entail 100 identical repetitions for him or her, the initial variations will be so large for the learners, that they will feel they are saying new versions for quite many times. But soon enough the chorus will approach and reach the same or almost the same pronunciation as the teacher has. The learners will feel that the ease of pronunciation increases dramatically during these exercises, and that the feeling of having it automatized will not appear until after a large number of "same" productions. But by this time the learner's feeling of success will be very strong, particularly if the teacher keeps praising the learners liberally. And generous praise and encouragement is required regardless of degree of achievement. The feeling of one's own success is a strongly addictive feeling, almost as addictive as a drug. As a result of the chorally facilitated automatization process coupled with teacher's praise, the motivation surges and soars, and the learners will want to do these exercises more and more, over and over again! The limiting factor will only be the teacher's patience ...

When, after some 25-30 repetitions, the learners have reached a consistently correct and acceptable pronunciation, only then does the exercise start. It should never ever be terminated at that point, because it is this correct pronunciation that is supposed to be wrought into the learners' procedural memories, not the wobbly way up to it. If the exercise is terminated before the correct pronunciation is automatized, then the whole exercise time may well be wasted for nil remaining avail. Perhaps this is what killed the ALM - too few repetitions?

For the same neurobiological reasons as outlined above, every instance of talking at all is an instance of speech "exercise" - including when the speaker has not been acceptably guided into the desired pronunciation of the target language. In effect, then, fossilization of a "foreign accent" will be a ruthlessly self-perpetuating process if the accent is left to remain. In the present author's experience, this is an unnecessary fate.

**Cornerstone #4: Categorical Perception**

On hearing the sounds, rhythms and melodies of our mother tongue(s), we are seldom aware of the infinite variations from person to person and from instance to instance of the "same" pronunciation. The speed of our hearing perception is due to the fact that we don't have to pick up and analyse the acoustical reality as such, but only identify our mentally and culturally conditioned categories, or "targets" of speech.

Categories exist on all levels: phonemes, morphemes, lexemes, syntactic structures, prefab phrases ("giant words" (Pawley & Syde, 1983)), sayings, proverbs, etc. By categorical perception we will hear selectively what the context suggests, sometimes even mis-hear due to where our thoughts happen to be at the moment. This is a
top-down process, the above-mentioned pattern recognition mechanism: We hear what we expect, based on previous knowledge. We can never perceive what we don't expect.

Compare how we label the colours of the rainbow. In many languages there are "seven colours", but physically, of course, there are millions of colours. It would be impractical to have to name all of them, even though we can see all of them. If we work in a haberdashery, we can probably name many more colours than the seven basic ones. All categorizations evolve in response to the particular needs and circumstances.

Problems may arise if two interlocutors are not agreed on the labels and, particularly, on the delimitations of categories. Colours are a typical example of culturally conditioned delimitations. What is the colour of the sky - blue? Of unripe fruit - green? Of the "go" traffic light - green? In Japanese, all three are labelled "ao". Of course, Japanese people can also see millions of colours and work in haberdasheries and appreciate the physical difference between the "blue" sky ("aozora") and the "green" light ("aoshingoo"). On learning English they just have to be informed of and trained on the new delimitations of categories and their labels. There is nothing magical about speech sounds or other linguistic units to make them functionally different from colours. We are just almost like deaf to the category boundaries of other languages.

The categories of our first language thus function like preconceptions, or even prejudices, of how they should sound. Or, rather, of what they should be. Because we don't hear physically with our ears but mentally with our brains, and we are almost like deaf (the above-mentioned amblyacusis) to the within-category variations. But thanks to that beneficial fact, most reductions in natural running speech go undetected. We are usually not aware of them. Also, we completely disregard such incredibly large acoustic differences as are due to speaker's age, sex, body size, mood, attitude, chewing-gum, tooth-ache, or whatever.

At the same time and for the same reasons, we are super-experts on detecting even the most minute transgression past the limits of our categories. This talent is the very "purpose" of categorical perception. It is essential for the correct perception of, e.g., ship/sheep, or eyes/ice, in running speech, whereas any within-category variations in the pronunciation of just 'eyes' will only add to the emotional or extralinguistic context. At least for the native speaker and listener ... Any unexpected or inadequate transgression past the category limits will be regarded as a joke if isolated (Crystal, 1998), or as another dialect or a foreign accent if consistent. Our categorical perception is an impregnable "goal keeper"!

The pattern recognition mechanism does not work very well if the rhythm of speech is very different from the expected. Prosodic patterns are also classified in (language-specific) categories. A recognizable pattern is required to activate the prediction machine. Or else the listener will have to analyze the details and invent probable patterns. Not only does this take longer time, as already mentioned, it also has to direct his attention more to the mistakes than to all that is correct. The speaker risks getting a lower estimation than deserved. A correct prosody, on the other hand, will camouflage most minor mistakes in vowels and consonants, thanks to the very same top-down category prediction processes in perception.

**Strategy #4: Find Category Limits, Not Centres**

Practice finding the limits of permissible variation, not bull's-eye pronunciation such as in traditional pronunciation manuals. That kind of "canonical" pronunciation is never relevant to natural speech. As in industrial quality control (QC) algorithms, alarms are only activated when measurements fall outside the tolerance limits more often than expected. The teacher should keep an alternation of chorus practice and individual practice tenaciously going on and on, enthusiastically praising every utterance, and, when deemed necessary according to his or her QC standards at the moment and for the purpose of that particular exercise, adding tricks to pull the learners back into the tolerance area. Aided by the consequent statistical effects (Cornerstone #3), the learners will gradually develop their feelings for where the category boundaries are, and thus acquire a categorical perception mechanism that has a real chance of becoming native-like.

**Cornerstone #5: Compensatory Articulation**

Compensatory articulation is the speaker's counterpart of the listener's categorical perception. Thus it will be practiced at the same time as in Strategy #4. Even when we mean to say the "same" thing, there will be some
differences every new time we say it. Humans have 146 muscles directly involved in the production of articulated speech, not counting the diaphragm and chest muscles for the air pressure, abdominal muscles for the support of air pressure, neck muscles for head posture, or limb muscles for gestures and body language. The balanced tensions and neural commands to all of these muscles must be perfectly coordinated millisecond by millisecond, but signal timing and intensity have to be very, very different depending on where the particular articulator just has been and whereto it is heading next, and also depending on head and trunk posture, amount of remaining air pressure in the lungs, etc.

Saying, for instance, "Piccadilly" looking up, down, to the left, or to the right, or with lungs full or half-empty, entails very different tissue tensions and therefore requires very different articulatory commands. Try it! The learners will have to practice this, too. It is deeply automatized in the native speaker, because of the audio-motor category templates stored in our procedural memory. Articulation thus is result-guided. We know how we want the sound to be, acoustically, and we can always get it there in some way or another. Try, for instance, to say both "ee" and "oo" with your lips firmly held in a wide smile. You will manage the unrounded "oo" quite well even without prior practice. Try to feel what you do with your tongue instead, when you can't use you lips in the usual way. Ordinarily, of course, the lip shape is the major contributor to the acoustics of "oo", but now the tongue compensates for the lack of lip movements.

That you, dear reader, managed the test in the last paragraph shows that you are, in effect, an expert on acoustical phonetics. Every native speaker is. The second-language learner has to start almost from scratch. Well, actually not scratch at all but rather quite near the target, because everyone will already know at least one language perfectly. That is indeed not an obstacle but a wonderful resource to depart from. Ready-made language tracks in the neural lawn! All languages are more equal than different (Baker, 2001; Trask, 1998). Just move the category limits slightly ...

**Strategy #5: Spend Minimal Effort for Just Sufficient Clarity**

The tricks for the compensatory articulation of the new language happen to be the same as in the previous four strategies: Use the ears, practice alternatingly in chorus and individually, practice full phrases, give priority to prosody, repeat sufficiently many times for a statistical mass effect, and concentrate not on canonical pronunciation but just on heeding the limits of permissible variation. This should lead to more relaxed, or "lazy" pronunciation, as for the native speaker who automatically spends minimal effort for just sufficient clarity. Include the whole body and any available multimodal gadgets. The teacher will have to point out salient details, one at a time. Use a QC approach to build compensatory articulatory competence. Immediate feedback (praise and corrective tricks) is essential. The average working memory lasts only a couple of seconds. Teacher/trainer feedback later than that will not aid much in the consolidation of relevant, new synapses.

The immediate auditory feedback from one's own voice and speech is also essential. It refines the co-development of the speech organ habits with the auditory system and the mirror neurons connecting them. Though the chorus helps the individual's articulation through the mirror neurons, one still hears one's own voice the strongest. These extensive repetitive exercises aim at establishing robust, auditory memories of the learner's own voice and own pronunciation when he or she manages to pronounce "perfectly". These "own" templates will work much better than any memory trace of the teacher's voice or cassette tapes. Let's say farewell to the "say-after-me" era and welcome the "say-in-unison-with-me" era.

**Epilogue**

As should be understood from the above, language acquisition is easier, and communication will be more fluent and easier for both speaker and listener, if only the prosody conforms to the local norms. If carried out well, the tedious choral work of the initial weeks will not have to be continued much longer, but it will turn out to have been a well-spent "waste-of-time". That is, not a waste of time at all. Within 6-8 weeks the motivated L2 beginner will have a chance to attain a performance level corresponding to, or exceeding, a five-year old L1 speaker, and a competence level much higher than that. What is 8 weeks compared with the alternative, having to speak with an unwanted accent for the rest of one's life?

The average L2 learner may have some remaining problems with certain sounds and more complicated grammatical phenomena. But he or she should have a native-like prosody and a well automatized grammatical...
base for the coming acquisition process, the so-called superstructure of the language, in a similar fashion as most native-speaking children will have it by the time they begin school. This is an ideal start, smartly evolved through millions of years for first-language speakers, and now emulated and adapted for modern second-language pedagogy. Good luck!

References


