

Which Phonological Model does Electromyographi Analysis Support: The Phoneme, the Feature, or the Articulatory Gesture?

メタデータ	言語: eng 出版者: 公開日: 2007-06-28 キーワード (Ja): キーワード (En): 作成者: Charles, Jannuzi, 吉澤, 正尹, Fawaz, Annaz, YOSHIZAWA, Masatada メールアドレス: 所属:
URL	http://hdl.handle.net/10098/755

Which Phonological Model does Electromyographic Analysis Support: The Phoneme, the Feature, or the Articulatory Gesture?

Charles Jannuzi, Masatada Yoshizawa, and Fawaz Annaz

(Received 14 September 2004)

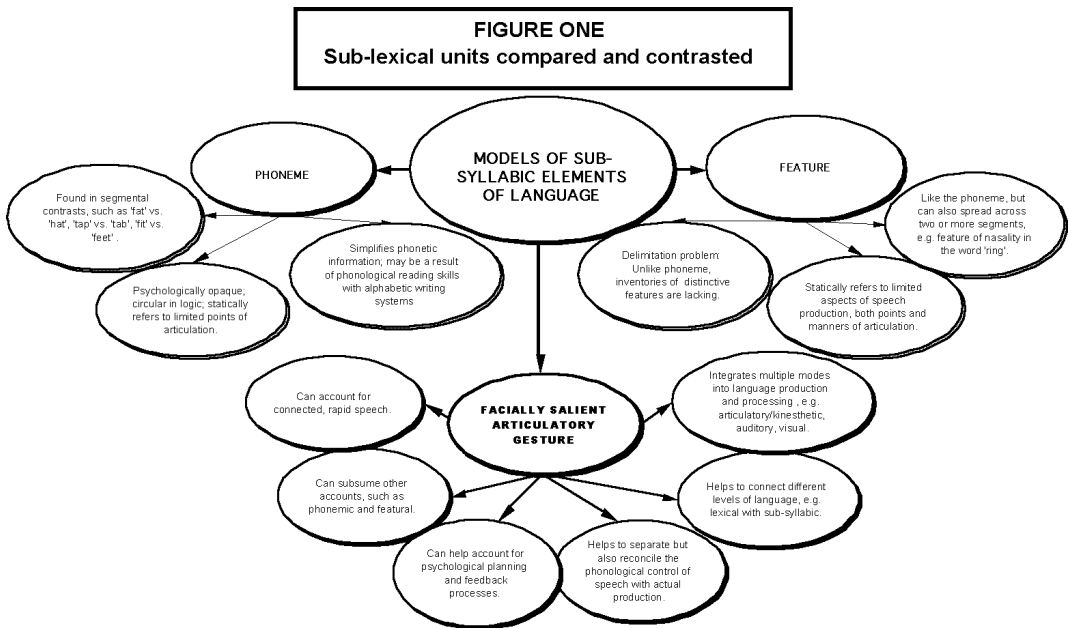
Introduction

This paper summarizes the analysis and interpretation of the results of two electromyographic procedures in experimental phonology. The results of electromyographic experiments have been interpreted and analyzed using concepts and theory from linguistics, applied linguistics, and phonology, specifically articulatory phonology. The first electromyographic procedure on one native speaker of English obtained data on the consonant sounds of English. The second electromyographic procedure was used to explore the large vowel system of English.

Based on the results of these experiments, we propose a new theory about the basic sub-lexical unit of speech production and perception. This paper posits a new, discrete, invariant, psychological unit of phonology that functions below the level of word meaning to organize language. This model is a variation of the articulatory gesture of articulatory phonology and phonetics, and it has implications and applications relevant in many areas of applied linguistics and language education, including native language arts, second and foreign language learning, and literacy. In order to contrast the new concept with the previously established concepts of the 'phoneme' and 'feature', we will call the new phonological prime the 'visual articulatory gesture', or, alternatively, it can be referred to as the 'facially salient articulatory gesture'. The advantage of this new basic sub-lexical unit in phonology is not merely the need in linguistics, applied linguistics and educational linguistics for an abstract model that makes better phonetic and psychological sense. Rather, we feel strongly that any model more true of linguistic and psychological reality will yield better concepts, principles and practices for the classroom and materials.

The theory that emerges from our research helps to solve the problem of the lack of phonetic realism that plagues structuralist, behaviorist and formalist accounts of the phonology of a language in actual acquisition and then communicative use (production and perception). In part, this model of phonology is based on a view of language as a learning system that builds up to a learned, stable state of functional complexity

(that is, the flow from language acquisition and learning to fluent use of a language to learn and communicate). The 'learning to learn' stage involves necessary and sufficient inputs and feedback from visual, acoustic-phonetic and kinesthetic signals. We call the most basic, sub-lexical, phonological unit of this model (and indeed all language use) the 'articulatory gesture'. However, unlike previously established conceptualizations of the term 'articulatory gesture', which never really address what is meant by the term 'gesture', our basic sub-lexical unit involves 'faciality' or 'facial salience' in the visual and physiological components. In this way we clarify why articulatory gestures are gestural in a linguistic sense and can help account for rapid, reduced, connected, co-articulated speech. Unlike the descriptively simplistic but non-explanatory abstractions of the phoneme or feature, articulatory gestures ARE NOT merely formalizations of repetitious, sequenced movements of articulators tracked at prominent points of articulation. Rather, the articulatory gesture as a unit of phonology helps models psychological control of both language production and perception. For a schematic overview of the articulatory gesture with the previously established analogues, see Figure One below:



Legacy concept - the phoneme

This term is perhaps most often defined and thought of in linguistics and language teaching as the smallest sound unit to create lexical contrasts in a language. For example, we might posit the existence of the /b/ phoneme in 'bat' or 'bin' if we contrast them with the rhyming words 'at' or 'in' and see that these words dif-

fer from the former by the absence or presence of one consonant phoneme. Or we might isolate a vowel contrast by placing 'bet' alongside 'bit', thereby helping us to distinguish between the vowels /e/ and /i/. There is something troubling, however, about the need for using words or lexical level meaning to help us define or determine what sub-lexical and even sub-syllabic sound segments are. Moreover, we have to think of phonemes as idealized or psychological categories of sounds, not actual instantiations of sound categories. This is done to the point where phonemes subsist as mentalist or social super-structuralist objects subsisting in some non-material realm, shorn entirely of their phonetic identities. Another aspect to consider is just where in words phonemes can occur--that is be instantiated. We might think of the nasal-velar sound at the end of the word 'ring' as an example of a phoneme of English, but the distribution rules for that phoneme in English determine that such a sound is not possible at the beginning of a syllable or word.

Unfortunately, overall, the 'phoneme' does not hold up to any close linguistic scrutiny of how languages are actually spoken, conveyed through the air as sound, received as audible material, and then perceived or integrated into linguistic understanding. Real speech doesn't naturally segment--we don't speak in discrete blips of Morse code. We can artificially segment spoken English, but the 'sounds' you will encounter will far exceed the 44-48 inventories phonemic accounts always give. The phoneme cannot be found in the mouth; it cannot be found in the air coming out of the mouth; it cannot be found in the air going to someone's ear; and it cannot be found in someone's ear. So then we are supposed to believe it is a socio-structuralist or psychologically real object, in which case we hardly need phonetic criteria to delimit it. And this is why phonetic analysis of phonemes always flounders on phonological nonsense or at least phonetic oversimplifications, if not all-out contradictions.

Take for example two of the most common types of sounds in English, indistinct, neutralized vowels of very low productive intensity (such as unstressed /i/ and schwa) or glottal consonants which are articulated at the extreme ends of the vocal tract (the glottis and front of the mouth). How should we phonemically interpret the schwa? Is it phonemically speaking the most common vowel sound of English and a category in its own right, or is it so common because it is an unstressed allophone of so many other vowels? Why should so many distinct vowels converge on the same sound for an unstressed allophone? Or what about geminate glottal consonants (such as the glottal /t/ we might find in the middle of the word cattail)? Phonemic accounts of the schwa or the glottal geminate might say that they are phonemes in their own right and that when they alternate with other sounds, these are processes of morphophonemic alternation. However, what if we said the schwa is just a phonetic variant of most of the vowels of English? And the glottal geminate a variation of English consonant stops. After all, there is enough phonetic similarity to make the case.

Other difficulties of interpretation and explanation abound. How should we phonemically interpret vow-

els in languages with diphthongs and triphthongs? How should we actually phonemically interpret the 'ng' sound(s) at the end of 'ring' or 'sing'? Native intuition is that they are two sounds or sub-syllabic elements, such as two concurrent, distinct but overlapping features (which is why no one has a real problem with the digraph of the orthography). But you could put these words into minimal pairs and come up with all sorts of contrasts. *Ring vs. rig, sing vs. sin*, etc. One phonemic account might make the sound fall under the same category as 'ng' because the 'ng' sound is the opposite of the 'ng' and comes only at the beginning of a word or syllable--except the failure to meet any criterion of phonetic similarity might be invoked. Or how should we treat the *in-/im-* prefix? Is the prefix of 'inert' and 'immobile' different in its forms because of morpho-phonemic variation or could one argue that either the /n/ or the phonetically similar /m/ is actually an allo-phonetic variation of a phoneme? The phonemic model for teaching and learning a foreign language's phonology predominates and is largely a formal model inherited from structuralism. Even if we supplement or supplant the idea of phonemic segments (segmentals) with suprasegmentals (e.g., intonation), the basic idea still centers learning pronunciation on the perception of arbitrary, social-systemic contrasts enabling an individual as language user or learner to understand spoken language.

However, it is impossible to locate the phoneme or contrastive segment in articulation, acoustics in physical space, or in reception and immediate analysis of the speech signal. This delimits it, if it exists at all, as a largely inaccessible and overly abstract, logical category taken away from actual speech and the psychological control of the vocal tract. Such opaque, black box concepts do not transfer well to the classroom, where effective simplification is necessary for teaching and presentation to have an impact on language learners. One might ask of the phoneme, if we do not say it and cannot naturally find it in the speech signal, why do we think it actually exists in language? It could be conjectured that the concept of the phoneme is actually an metalinguistic artifact of psychological perception--a super category imposed on sounds and the vocal tract--stemming from linguistic insights about meaningful language use or even literacy in an alphabetic language. We could also argue that some sort of concept of the phoneme is a convenient fiction which allows us to refer more consistently to key points and manners of articulation in written language than does standard English spelling, which is more geared toward preserving etymological relationships across inflected and derived forms.

Legacy Concept - the Feature

Early on in structuralist approaches to phonology (and then later, transformational ones as well) another idea was posited that supplemented in more detail the earlier concept of the contrastive phoneme--that of the contrastive feature. Phonemes, it was theorized, could be broken down further into distinctive features. For example, whether or not they are phonemically contrastive, phonetically speaking, what typically sepa-

rates a language's /t/ from its /d/ is the feature of *vocalization*. Or, another example is how voice and a lack of aspiration might distinguish an initial /p/ from an initial /b/. One difficulty of breaking phonemes down into features, however, is that the aspects of speech that have been called features are a confusing mix of psychological, articulatory-gestural, phonetic and acoustic phenomena. Typical notions of features move back and forth between articulatory criteria (a point or manner of articulation in the vocal tract or respiratory tract or an acoustic effect found on an oscilloscope). Is something a feature only if a listener hears it or could a feature be something that is physiologically experienced and subsequently anticipated by the speaker? A second problem is that, as described in much discourse, they are not truly sub-syllabic; at least phonetically and acoustically speaking, features demonstrably spread over whole syllables, words and even word boundaries. Features, then, if we actually break up speech in order to demonstrate their existence, are supposed to work more like the various notes of a chord either struck almost simultaneously or plucked out in quick succession but sustained and stretched over an entire bar (in this case syllables and syllable sequences) to create harmonies and dissonance.

Evolution of language as gestural in nature

The human ability socially to convey thoughts, intentions, emotions, beliefs, and culturally bound ways of living largely depends on the structured use of language. This cognitively controlled, structured system for communication, we contend, evolved first as a visual-gestural system of body language quite analogous to the sign language of the deaf in use today. That is, we are talking about a gestural language that involves not only the hands and arms, but also movements of the muscles of the face to produce a form of controlled speech that is more reliant on the visual conveyance of information than the acoustic mode. The full development of human language as we now know it, however, overlapped with the emergence of considerable auditory and phonetic abilities crucial to the survival of the human species. These beneficial auditory and phonetic talents also took on communicative functions contributing to the survival and adaptation of the species.

Over time the visual-gestural system of language converged with the auditory and phonetic powers to produce what we know today as the human language facility. It might make more sense to view the auditory and phonetic aspects of human language as dominant over the visual and gestural ones. Also, not all visual-gestural aspects of communication are linguistic in nature, though many can be specific to particular groups and cultures. However, it might well be the case that visual and gestural abilities are still more integral to the psychology of language control and acquisition. For example, the use of gesture is two-part. It provides a visual signal for someone at the receiving end, but the person producing the gesture also experiences it physiologically.

The ability to communicate with a human language depends essentially on a psychologically controlled, coordinated speech and auditory system for the planned production and meaningful perception of language. It should be pointed out that speech production itself depends on a convergence of more basic systems, such as the ones, which hear, breathe, eat, and make non-linguistic noise. And hearing has a more basic non-linguistic role enabling humans to distinguish and make a phonetically diverse set of noises for communication, such as signaling and sound camouflage. Human language, however, has more essential aspects to it than speaking and hearing perception. It also involves visual and kinesthetic aspects and structural complexity that ranges from the phonological to the lexical to the syntactical. The visual and kinesthetic aspects of phonology, however, quite likely play an even larger role in language acquisition than they do in mature, fluent, native language use for everyday communication.

A new phonological prime proposed

In lieu of the previously established concepts of the phoneme and feature, we call the most basic, sub-lexical, phonological unit of speech production and processing the 'articulatory gesture'. However, unlike previous conceptualizations of this term (for example, Browman & Goldstein, 1992), our basic sub-lexical unit involves 'faciality' or 'facial salience' in order to explain how a unit of speech can function as a linguistic 'gesture'. It must be noted here, though, that which level of language should be used to interpret speech production and processing remains a theoretically undecided issue. Does the articulatory gesture map onto language at a sub-lexical level (such as the syllable or mora)? Or does the articulatory gesture actually correspond in an explanatory manner to the spoken and psychological level of word meaning--that is, the allomorph and morpheme? If the reality of the latter case holds, then morpho-phonology would assume primary importance in any research program. Clearly more conceptual, theoretical and experimental undertakings are required for this issue to begin to be resolved.

Using linguistic analysis and interpretation of the results of two experiments in electromyography, we propose a new theory of phonology concerning the basic unit of sub-lexical language. Modern phonology has long sought a basic, psychological, discrete, invariable unit of language subsisting beneath a word level in order to closely model, describe and explain language acquisition, processing, perception and expression. That is, in order to solve the age-old problem of 'how infinite use is made of finite means', phonological inquiry needs a basic, stable, sub-lexical unit that works across all aspects of a language and across all speakers of a language. Such a unit is not only deduced to exist because speech can be segmented into consonants and vowels that form syllables. This could simply reflect a phonetic reality of speech that has been analyzed by linguists. Rather, a phonological prime subsisting at a sub-lexical level of language must also function as a unit in the mental language planning stage that controls meaningful language use.

Some Implications for Applied Linguistics and Educational Linguistics

Such a psychologically, physiologically and phonetically realistic basic unit for phonology should yield better teaching and learning materials (including software) in applied, practical and clinical areas such the following: (1) foreign language teaching and learning, (2) speech therapy, and (3) learning disabilities, such as reading and text processing disorders. This approach should also have important implications for the development of speech recognition for automated word processing, language translation and artificial intelligence.

There are available in applied linguistics various approaches to studying, describing, analyzing and explaining the production, transmission and perception of a language's phonology. However, one crucial problem is turning it into useful information for second or foreign language pedagogy. Phonological concepts and terminology for teaching often seem overly complex and abstract--if not outright contradictory--to both teachers and students. One possible reason for this perceived difficulty is that, in fact, many of the terms and models used to teach phonology simply are not useful for adolescents and adults learning a phonology. On the one hand, the meaning of terms in phonological discourse comes to seem opaque to students and even the teachers attempting to explain and demonstrate them. On the other hand, the concepts are too simplistic and static to do justice to the phonological, phonetic and physiological complexity that a learner must deal with in mastering a second or foreign language's phonology.

The articulatory gesture and its implications for language learning

A third approach to FL pronunciation and phonology would be to appeal broadly but coherently to those aspects of speech that apply to the phonetics, physiology as well as psychology of speech. Rather than being determined through analysis of static, binary contrasts, the sub-syllabic units of speech are deduced to exist and represented through dynamic descriptions of a complex of movements occurring in the vocal tract, mouth and facial muscles. This is called an articulatory-gestural approach or articulatory phonology. According to this approach, the basic units of phonological contrast are gestures, which are also abstract characterizations of articulatory events, each with an intrinsic time or duration. Utterances are modeled as organized patterns (constellations) of features, in which gestural units may overlap in time. The phonological structures defined in this way provide a set of articulatorily based natural classes. Moreover, the patterns of overlapping organization can be used to specify important aspects of the phonological structure of particular languages, and to account, in a coherent way and general way, for a variety of different types of phonological variation. Such variation includes allophonic variation and fluent speech alternations, as well as 'coarticulation' and speech errors. Finally, it is suggested that the gestural approach clarifies our under-

standing of phonological development, by positing that prelinguistic units of action are harnessed into (gestural) phonological structures through differentiation and coordination. (Browman & Goldstein, 1992, p. 155)

Although not well known or understood in FLT and FLL, an articulatory-gestural approach to phonology (or articulatory phonology) may well hold out the most promise for reuniting pronunciation practice with communicative language teaching and learning.

One problem with any theory that seeks to explain how language is spoken because of what the tongue wants (ease of articulation) is that it might not take into account what the ears easily hear. A language user's vocal tract that has to repeat itself with emphasis actually ultimately does more work. Spoken language as a system built on give-and-take communication is pushed and pulled between the needs of the speaker and the listener (just as writing systems have to fit the needs of those who write and those who read). Going toward language that is rather indistinct, lacks intensity and overlaps sounds (co-articulated segments, super-syllabic features, reduced vowels, glottal consonants) might make it faster for the speaker, but this only has to be optimized to the level of how fast a listener (as language user) can take it in (which has physical limits). A rate of output beyond the point of what a listener can perceive does not contribute to the efficiency of production or reception since it would cause a breakdown in communication.

Different languages, dialects, and accents arrive at different sets/constellations of articulatory gestures (or articulatory routines) to get the job done. If there is considerable overlap of grammar and lexicon, then mutually intelligible forms of languages can exist, despite quite a bit of variation in how things are pronounced. Any number of ways could arrive at basically the same speed of output for optimum reception and would be well below our maximum speed of output if our lazy tongues ruled our heads. But what would be the point of being able to speak so fast and indistinctly that no one could understand you?

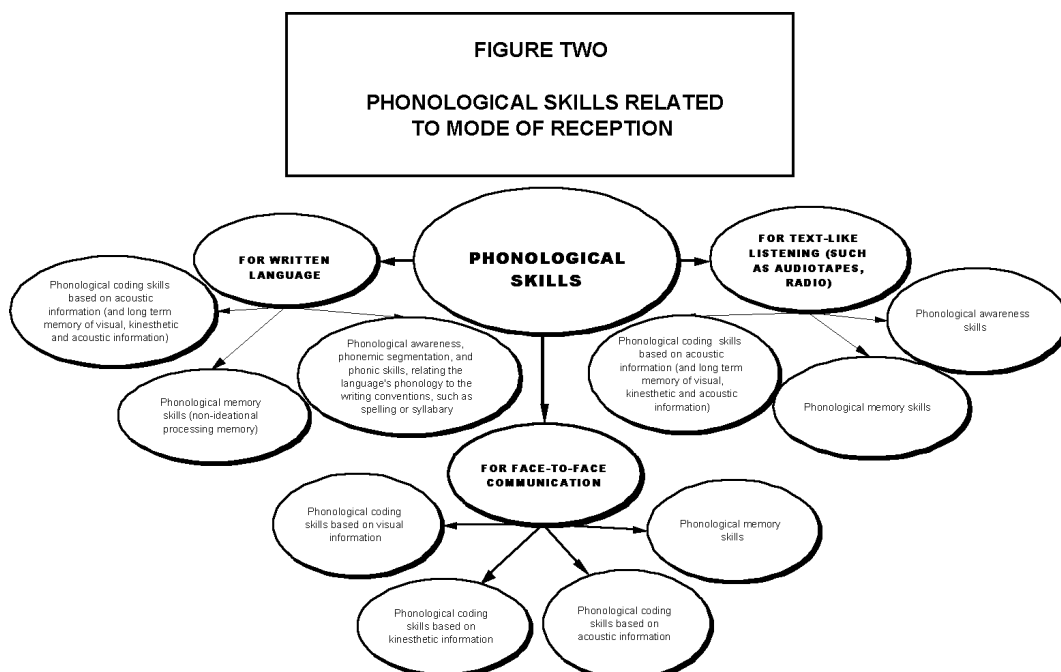
A facially prominent account of articulatory gestures

Following the example established with the electromyographic analysis and pedagogical recommendations of Koyama, Okamoto, Yoshizawa, and Kumamoto (1976), we propose to take prior accounts of the articulatory gesture and modify and simplify their focus for the purposes of L2 pedagogy and learning. The rationale for this is, in part, based on our understanding of both language evolution and natural language development in individuals. One possible way to account for the human ability to make meaning systematically is to see the human vocal ability with language as a fortuitous adaptation of our respiratory, upper digestive and auditory tracts that extends our ability to gesture semiotically. The face, however, is a transitional area that serves a role both in the vocal apparatus and in the purely visual-gestural system. Indeed, with the face's and mouth's exterior as an interface or transitional zone, it could be said that the vocal appa-

ratus and the visual gestural tools of the upper body form a seamless semiotic continuum.

One clear advantage of an articulatory gestural account of phonology is that it gives a dynamic, physiological basis to our ability to use a language to communicate. Moreover, such an approach might also help us to account more holistically for the ability to handle fast (i.e., normal), reduced, co-articulated connected speech in everyday spoken communication. Not only do we hear such speech, but our prior physiological experience of language use helps us to anticipate and fill-in information missed from the audible portion of the stream of speech.

An articulatory gestural account that focuses on the face and the mouth most vitally allows us to reconcile the natural, untutored, pre-literate language development of a native speaker with the possible course an L2 learner would be better off following. Consider, native language acquisition depends crucially on both auditory and visual inputs and feedback from caregivers. Even if fluent, stable language ability in humans has shifted heavily toward the auditory part of the semiotic continuum, it seems most likely that visual input (in coordination with the stream of speech) from the faces of immediate caregivers provides necessary types of both input and feedback to infants acquiring a language. Note just how an infant must experience language and its development: the infant experiences making movements in its own vocal tract and face; s/he feels and hears the sounds thus produced directly through the medium of the head; s/he hears the sounds going through the air and back into the head by way of the ear; s/he hears the caregiver respond (often in exaggerated and simplified adult speech); s/he most crucially sees in three dimensions the facial



and upper body movements of the caregiver. No idealized schematics of the interior of the vocal tract of either the infant or the caregiver are required, nor is a visual perspective on the inside of any human mouth necessary. For a schematic overview that relates possible phonological units with type of interaction and/or mode of reception, see Figure Two (above).

What is electromyography?

Electromyography is a means to measure and graphically record in controlled settings the electrical activity of muscles, including, of course, the muscles used in producing speech. Muscles generate electric current when contracting or when the controlling nerves have been stimulated. Electrodes usually attached to an abraded area of the skin over the muscle pick up the impulses. The output of the muscles can then be displayed as wavelike forms on an oscilloscope and recorded as an electromyogram (EMG). The audible signals which stem from the activity of the vocal tract can also be recorded simultaneously, though it must be remembered that this audible stream of speech is an acoustic realization that results from the underlying psychological cognition, including sub-lexical, sub-syllabic manipulation of phonological units into larger structures of language (even if speech control is experienced at a point of subconscious control).

Language conceptualization and language planning, as cognitive processes, causally precede, but also overlap with speech production. However, the relationship in actual speech performance is a complex one: self-monitoring of speech (both acoustic and articulatory) as well as visual and acoustic feedback from an interlocutor can to alter or reinforce planned speech, which then affects the subsequent articulatory performance of the speaker.

Electromyographic techniques could be used to measure activity all through the internal parts of the vocal tract; however, such applications would prove impractical without very invasive--even surgical--placement of the electrodes. Moreover, once in place, the set up would interfere with normal speech. After Koyama et al. (1976), we instead propose the application of completely external, facial electromyographic techniques. This is because we are looking for the sort of common, salient visual and kinesthetic experiences, inputs and feedback that might naturally guide young language learners in their phonological development as an integral part of the greater category called language acquisition. In other words, we propose the use of electromyography as a means to better grasp, analyze and present what is most invariant and perhaps even holistically essential about phonological development in such a way that these insights can be applied to L2 teaching and learning.

Our data collection efforts and what the results reveal

Our data collection efforts are still somewhat preliminary and have involved only one subject (an American native speaker of English, one of the authors, Jannuzi). We have collected extensive data sets on both the vowels and the consonants of English, and would next like to generalize this to a larger group of native speakers of English. Moreover, in the future we plan to correlate visual and audio materials systematically with electromyographic data so as to triangulate the physiological-kinesthetic elements of controlled language production and perception with the concurrent visual and acoustic phenomena. However, in presenting our conceptual and theoretical work here, we also have the experimental and pedagogical insights of Koyama et al. to draw on. They have already shown, using electromyographic data and photographs of the face, that there are specific but regular ways of using facial muscles in pronouncing the English consonants. Moreover, they demonstrated how electromyographic data generated during speech can be used to isolate points of instruction so that teachers can better train Japanese EFL learners in pronunciation and phonological development.

We have already been able to come to some tentative but interesting conclusions about the possible physiological and articulatory gestural aspects of phonology. For example, a phonemic account of English might place /l/ alongside /r/ as an important contrast that an English speaker has to make. However, teachers must ask if saying there is a single contrast is actually very useful in order to teach students how to make the sounds during actual communication. Acoustically speaking, English /l/s and /r/s produced in some environments can appear to sound quite similar and hard to distinguish (perhaps because of three-formant, voiced aspects, which make both /l/ and /r/ much like vowels). Phonetic or featural differentiation, when it hits upon points of articulation, starts to be somewhat more useful. It is usually taught that an English /l/ is an alveolar lateral whereas the /r/ is post-alveolar or retroflex. But both the /l/ and /r/ in actual language display an enormous, confusing range of variation.

Can an articulatory gestural account focused on the facial muscles involved in speech (namely, *M. temporalis*, *M. masseter*, *M. levator labii superioris alaeque nasi*, *M. orbicularis oris*, *M. depressor labii inferioris*, *M. digastricus venter anterior*) help to differentiate what are syllables or words with /r/ sounds from those with /l/? Our initial conclusion is, yes it could. A very preliminary exploration of these two sounds focusing on the muscles of the face indicates that there is clear, visible differentiation patterns to be found across /r/s and /l/s. Most significantly, a syllable or word that begins with an /r/ sound is articulatorily prepositioned to a mouth shape somewhat like an English /w/ or the vowel /u/, no matter what the following vowel that forms the nucleus of the syllable is. On the other hand, in terms of facial movement and anticipatory shaping of the mouth, the /l/ is far different. Visual investigation reveals that typically the /l/ coarticulates with the following vowel; in articulatory-gestural terms, we could say that the vowel that is sup-

posed to follow the /l/ is articulatorily anticipated before the /l/ is actually made. We plan to explore this sort of physiological patterning in much greater detail, with focuses on the English /l/, /r/ and English's rather large, difficult sets of affricates and fricatives, which are problematic for learners of various native language backgrounds. Also, careful analysis of actual electromyographic data has revealed other patterns that, while not necessarily contradicting phonemic or featural accounts, can be used to supplement and clarify them.

In brief, here are some of the more startling aspects to language that electromyography has revealed:

1. Despite what traditional theory says, no phonemic or featural distinctions are singular differences. For example, a phonemic account of the English /l/ and /r/ sounds would say that the contrast rests on the difference of a single phoneme. That is, the /r/ in the word 'ray' is very similar, acoustically speaking, with the word /l/ in 'lay'; however, /r/ has the added 'feature' of retroflexion. However, electromyographic analysis reveals far more detail. English /r/ is more like English /w/, but can be differentiated from /w/ in terms of muscular activity, timing, and a slight difference in the shape of the mouth. Also, English /r/ also forms relationships with preceding vowels when it closes a syllable, while English /w/ only acts as the onset of a phonological syllable, not its coda. For example, English /l/, in terms of muscular activity, is much more like the English /d/, except in the visually perceivable aspect of timing--that is, an /l/ sound lasts longer than a /d/, and this difference can be found in the electromyographic data as well as visually perceived on the face of the speaker. In the position of the end of a word or syllable, English /l/ might also alter its gestural form through a relation with a preceding vowel sound.
2. Electromyographic data give tantalizing, psychologically significant hints as to the language planning stage of speech production, which falls in between conceptualization and actual speech production. In fact, electromyographic data reveal direct evidence of the physiological control that both precedes and accompanies actual speech production. One counterintuitive aspect thus revealed concerns the intuitive notion of speech being a sequence of sounds. In terms of the muscular activity that precedes speech production, we cannot say that speech is, phonologically speaking, a simple sequence of sound segments. For example, a one-syllable word ending in a /p/ consonant might display more muscular activity before even the first segment of the one-syllable word has been produced as sound. If we contrast the words 'mop' and 'mob', we see that the word-final /p/ sound is signaled in terms of muscular activity even before the initial /m/ has been produced. In other words, in terms of muscular energy used even before the word is uttered, the initial /m/ of 'mop' displays a higher energy level than the initial /m/ of 'mob' which could only be causally accounted for by the effect of having to plan for the pronunciation of a final /p/ instead of a final /b/.
3. As the example in number 2 above shows, the electromyographic data offers indirect evidence of a physiological interface between language planning and actual speech production. However, it is not clear at what level of language we can say that the articulatory gesture subsists. On the one hand, it would seem to

be a logical and useful sub-lexical unit of phonology that can subsume more static and incomplete models, such as the phoneme or feature. On the other hand, it might more closely match up with the unit of language known as a syllable. Or, more startlingly, it might be that, in connected speech, the articulatory gesture as a unit actually coincides with words and lexical phrases. Certainly the manifold differences that the electromyographic data reveal could be used to support the argument that one articulatory gesture equals one spoken syllable type or even one word.

Phonological coding ability

From theoretical and experimental standpoints, we argue that a facially salient articulatory gesture is the best model of psycholinguistically controlled speech at a sub-lexical, sub-syllabic level. However, within a more comprehensive view of language and literacy, there still might be a place for the concepts of phonemes and features. Certainly there are yet still more conceptual areas we must look at before we can begin to account for phonology in language acquisition, language learning, listening outside of face-to-face interaction, and literacy development. First, there are cognitive, linguistics skills called phonological coding (or processing) ability (PCA), centered on:

- Phonological perception and interpretation of phonetic or phonetically graphic data,
- Analysis/decoding of acoustic (and/or visual, graphic) signals in oral communication (and/or written discourse),
- Re-coding/encoding of linguistic input for lexical access and word recognition,
- Re-coding/encoding of linguistic input for comprehension and meaning making,
- Retention of language representations in short-term working memory (more specifically, phonological memory), and the linking of ALL these preceding points with long-term memory input, storage and retrieval--which is what makes phonological coding skills central to language learning, since phonology must be manipulated and stored as units such as features, phonemes, mora, syllable types, articulatory gestures/gestural routines and words (lexical units).

Metacognition: Awareness Skills in Language and Literacy Development

It is now a fairly well disseminated idea that language awareness and short-term memory skills at a phonological level play an integral role in literacy development in languages with alphabetic orthographies (Elkonin, 1963; Liberman, 1973; Liberman, Shankweiler, Fischer & Carter, 1974; Williams, 1995; Nation & Hulme, 1997; Stahl, Duffey-Hester, & Stahl, 1998; from a cross-linguistic perspective using pure research techniques, see Koda, 1987, 1998; for an ESL perspective using applied research case studies, see Birch, 1998). These skills are thought to comprise a metacognitive type of analytic ability which over-lay-

ers verbal language processing but remains separable from what have traditionally been called phonics skills, the latter of which emerge as part of reading in an alphabetically language. Thus, it is thought that phonological awareness skills follow from the phonological processing, production and perception skills that develop as a result of native language acquisition. However, they precede the development of phonics skills and beginning literacy and may play some sort of causal role in reading development. The related concepts of phonological and/or phonemic awareness are not well established within foreign language education, coming as they do mostly from theorizing about and research on native literacy in languages that are written alphabetically.

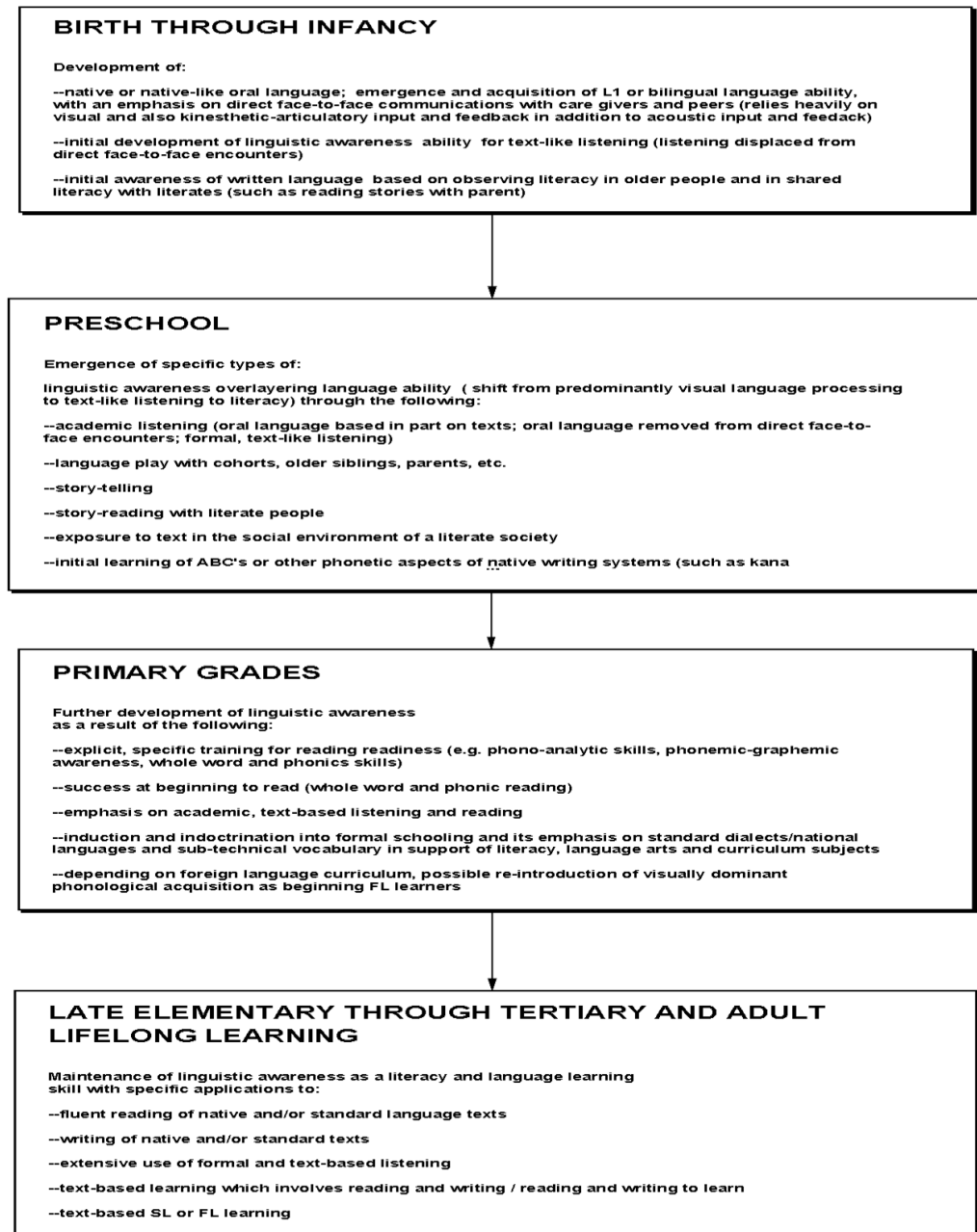
Epistemologically speaking, phonological and phonemic awareness abilities would seem to subsist somewhere between what Skehan (1998) calls 'phonological coding ability' and what have been traditionally termed phonics skills in native language arts. Phonics skills only come into play when alphabetic or syllabic writing conventions are associated with and/or analyzed into some sort of phonological equivalent during the reading and writing of text. In the case of reading written English, single letters and letter combinations functioning as graphemes (units of writing corresponding to units of sound) would be made to stand for single sounds and sound combinations in some sort of psycholinguistic process during reading; these representations might then be related to the phonology of spoken English to facilitate lexical access, which would then lead to the integration of lexical meaning into syntax and discourse. Phonological awareness skills may serve as some sort of metacognitive bridge between oral and written language processing.

It is often asserted that phonological awareness/metaphonological skills emerge before and may even causally underlie beginning literacy (hence the need to distinguish them from what has been called traditionally phonics). It might also be argued that this ability to manipulate an internalized language phono-analytically leads to the acquisition of phonics skills at decoding and manipulating alphabetic writing--especially if phonics skills are a key part of beginning literacy development and subsequent functional literacy. Phonological awareness skills are thought to be activated as a sub-component of the reading process because they help a reader (as language user) to decode and reconstruct information sampled from an alphabetically written text and relate it at one specific level with the reader's internalized phonology of the language being read. Such a step may be especially important in developmental literacy.

There is a different view, however, in which phonological awareness skills are seen as a fairly spontaneous development bridging native language acquisition of phonology with literacy development. This might undercut the hypothetical predictive, explanatory and instructional value of phonological awareness in reading instruction, since this view would make such skills appear to be more a result of success at beginning literacy than a causative factor underlying it. Another vexed issue is the orthography of English itself; although written alphabetically, English violates the alphabetic principle (one symbol=one sound) severely in numerous ways, to an extent that the reality of phonological and phonic reading of it has to be

FIGURE 3.

A HIGHLY SCHEMATIC OVERVIEW OF PHONOLOGY AS A PART OF NORMAL LANGUAGE DEVELOPMENT



drastically circumscribed, if not placed entirely in doubt. The language levels at which it can be said the code of written English is stable and determines the language read would be at the word level and above. For a highly adumbrated overview of the stages of phonological development in a literate society, see Figure Three (above).

Conclusion

There are various approaches to account for the production, transmission and perception of a language's phonology. However linguistically interesting such accounts and approaches are, how much useful information do they provide to teachers and students? One might conclude that many of the terms and models used to teach a second or foreign language phonology simply are not useful for adolescents and adults learning an L2s phonology, and, what is worse, they might be confusing to the extent they hold back learning. The problem might not be just one of technical complexity. The meaning of terms in phonological discourse may be too opaque and unnatural to students and even teachers. But however technical, the concepts could be too simplistic and static to do justice to the phonological, phonetic and physiological complexity that an L2 learner must deal with in mastering second or foreign language phonology. We have described and explained facial electromyography in support of a simplified gestural model of phonology. The electromyographic techniques we propose not only gives direct evidence in support of a gestural model, we argue it also considerable potential for the pedagogy of FL phonology in terms of teacher training and materials development (such as learning feedback software). What remains to be done must follow two major courses. First, we plan to expand our use of data gathering with electromyography to include a larger group of English native speakers, including a variety of dialects and accents. Electromyography will also be used to explore how to give useful and specific feedback to Japanese speakers learning English pronunciation. The second major track that we will pursue is the development of improved teaching techniques and learning materials that take advantage of the improved models and concepts of phonology that we have explained here.

References

- Birch, B. (1998). Nurturing bottom-up reading strategies, too. *TESOL Journal* 7(6), 18-23.
- Browman, C. P., & Goldstein, L. (1992). Articulatory phonology: An overview. *Phonetica*, 49, 155-180.
- Elkonin, D. B. (1963). The psychology of mastering the elements of reading. In B. Simon & J. Simon (Eds.), *Educational psychology in the U.S.S.R.* (pp. 165-179). New York: Routledge.
- Harris, T. L., & Hodges, R. E. (Eds.)(1995). *The literacy dictionary: The vocabulary of reading and*

writing. Newark, DE: International Reading Association.

Koyama, S., T. Okamoto, M. Yoshizawa & M. Kumamoto (1976). An electromyographic study on training to pronounce English consonants unfamiliar to the Japanese. *Journal of Human Ergology*, 5, 51-60.

Koda, K. (1987). Cognitive strategy transfer in second language reading. In J. Devine, P.L. Carrell, & D.E.

Eskey (Eds.), *Research in English as a Second Language* (pp. 127-144). Washington, D.C.: TESOL.

Koda K.(1998). The role of phonemic awareness in Second Language reading. *Second Language Research (London)*, 14(2), 194-215.

Liberman, I. Y.(1973). Segmentation of the spoken word and reading acquisition. *Bulletin of the Orton Society* 23, 65-77.

Liberman, I.Y., Shankweiler, D., Fischer, W.M., & Carter, B. (1974). Explicit syllable and phoneme segmentation in the young child. *Journal of Experimental Child Psychology* 18, 201-212.

Nation, K. & Hulme, C. (1997). Phonemic segmentation, not onset-rime segmentation, predicts early reading and spelling skills. *Reading Research Quarterly* 32, 154-167.

Skehan, P. (1998). *A cognitive approach to language learning*. Oxford: Oxford University Press.

Williams, J. (1995). Phonemic awareness. In T. L. Harris & R.E. Hodges (Eds.), *The literacy dictionary*, (pp. 185-186). Newark, DE: International Reading Association.