The development of sound systems in human language

K. J. Kohler
IPDS Kiel, Germany

Introduction

This paper presents a few ideas on the question of what is a speech sound, and it takes as its point of departure the seminal work by Lindblom and coauthors, especially [1,2,3,4,5,6,7]. So the aim is a contribution to the explanation of fundamental general phonetic categories in human speech and of descriptive language facts under the perspectives of universals in phylogenetic evolution, ontogenetic acquisition, communicative function and historical change. The following theoretical premises of Lindblom are essential for this discussion:

(1) The approach is deductive and hypothesis-driven, i.e. it is grounded in facts and assumptions that lie outside the field to be explained, in particular in the biology of the sound producing, receiving and understanding systems of homo loquens. (This does not negate the additional importance of sociological factors, but they will be excluded from consideration here.)

(2) Sound systems in human languages are the way they are as the result of balancing sufficient acoustic/auditory contrast for a listener and articulatory economy for a speaker under varying conditions of speech communication: they both tend to be maximised in the pursuit of communicative goals (H&H theory).

(3) Basic and elaborated or complex sound types have to be distinguished: [a] and [t] are intuitively simpler than [ɨ̃] and [ʛ], the task is to provide externally motivated reasons for this plausibility.

I shall attempt to apply principle (1) more rigorously in connection with (3), where, I think, under the influence of Maddieson's data-oriented approach Lindblom and Maddieson [7] come close to circularity because the postulation of their basic consonant set of 11 obstruents and 7 sonorants is primarily governed by the observed systems in the languages of the world; what we want is independent articulatory and perceptual justification for the notion of simplicity in elementary sound types. I shall also argue in favour of a more clearly two-stage evolution model - sound space delimitation followed by selection. The first step – of delimiting the human speech sound space and of constraining vowel or consonant systems – is considered to be grounded in articulation, i.e. in the neuromuscular and biomechanical constraints imposed by the anatomy and physiology of the apparatus available to humans for noise generation and in the general tendency to minimize articulatory effort. An acoustic/auditory selection principle then shapes speech as an instrument of communication between a speaker and a listener in a second step. This means that I wish to put a stronger emphasis on the articulatory basis of sound systems than Lindblom has done, due, no doubt, to his work on vowels having originated in the field of perception, although he also clearly recognised the importance of an articulation component in the shaping of vowel systems [4].

I shall put forth a set of ten assumptions and discuss the corollaries of each one, proceeding from articulatory considerations to perceptual constraints.

Assumption 1: the vegetative function origin

We can start from two sets of facts:
(a) Vocal language developed to address the ear as a more distal system of communication than the gestures of sign language, which have to rely on eye contact in face to face interactions under favourable lighting conditions. This development at the same time freed the hands and arms for other activities, an important factor in the struggle for survival. Seen from this perspective we can expect that sounds produced with the help of hands and fingers to shape the vocal tract do not form part of human vocal language at its original evolutionary stage. This rules out a large proportion of Fred Newman's "mouthsounds" [8] as possible speech sounds. The development of the whistled language of La Gomera [9] is no contradiction, because here the finger-shaped labial whistle is quite clearly superimposed on articulatory tongue movements of the primary vocal language, which was thus subsequently adapted for special long-distance communication.

(b) All sound-producing structures used in human language have primary vegetative functions that guarantee survival: the chest/lungs, larynx, velum, tongue, teeth, lips, mandible for breathing, protection of the breathing system, stabilization of the rib cage, swallowing, sucking, chewing. Speech is thus an adaptation of the preexistent anatomy and physiology of the vegetative system to the new function of speech communication to increase the chances of survival. From this fact we may conclude that only those noise productions enter into the formation of speech sounds that are related to the basic mechanisms already used for these non-speech functions. This excludes more of Newman's "mouthsounds" and all of Pike's minor air-stream mechanisms [10], i.e. all sound productions that employ generators which are not part of the vegetative basis. It leaves us with pulmonic, glottalic and oral air-stream mechanisms, with the exclusion of the egressive oral one.

Viewed from this angle clicks, although very rare in the world's languages [11], must be regarded as being among the basic archetypal phonemic elements of sound systems, rather than in Lindblom and Maddieson's elaborated set. This point is strengthened by the fact that outside the area of click languages there is, in Africa, a wide geographical distribution of labio-velars, which can be derived from (bilabial) clicks by eliminating the sucking motion and thus integrating the production into an overall pulmonic air-stream, while still preserving the tongue-lip configuration of clicks. Oral air-streams offer the additional advantage of not interfering with vegetative breathing, which may be very important in strenuous outdoor activity. Furthermore, they can have a high degree of perceptual salience, e.g. the palatoalveolar click. But these advantages are completely outweighed by the extreme limitation of paradigmatic differentiation as well as syntagmatic concatenation in oral air-stream productions. They have a restricted place continuum and very limited possibilities for manner categories (stop and affricate types, compared with the distinction of plosives, fricatives, affricates, nasals, approximants, trills, taps and flaps in the pulmonic mechanism); they thus lack the generation of vowel types, which are essential as carriers of syllable units (see Assumption 2). Moreover, the glottal and nasal modifications of clicks, which may be very elaborate in some languages (see !Xũ in [11]) are pulmonic by definition. The pulmonic mechanism is therefore by far superior, and selection has led to its dominance, relegating clicks to a peripheral position structurally and geographically, mostly in a paralinguistic function, rather than as linguistic speech sounds.

The modification of the pulmonic air-stream from vegetative breathing to sound production also shows the results of the adaptation and selection principles. The egressive air-stream becomes highly controlled and thus allows speech generation over long periods of time without the need of breathing in. On the other hand, the ingressive air-stream cannot be so regulated for physiological reasons, and is consequently only used paralinguistically, not for concatenative elements of phonological sound systems.

The vegetative origin of speech is also suggested by the way humans produce hesitation 'particles' in the speech stream, largely irrespective of the particular language they speak: when they run out of what to say and search for words, they use a non-word vocalization instead. The vehicle for this is an unrestricted oral or nasal air-stream, or a combination of the two, either simultaneously or in sequence, e.g. [ɛː], [mː], [ɛ̃ː], [ɛːm], [ɛ̃ːm]. In this the mouth opening (degree and rounding) is of secondary importance, and may vary a great deal from speaker to speaker, but it is interesting and significant to note that only more open-type vowel qualities, approximately representable in IPA symbolization as [ɛː], [æː], [aː], [ɜː], [ɐː], [œː], not [iː], [y:] or [u:], occur, and that the labial nasal is preferred. Moreover the vowel qualities of these hesitations do not coincide with those of the phonemes in lexical items of the particular language [12]; they just act as non-close vocal tract carriers and thus still retain many characteristics of the egressive phase of vegetative breathing. This leads us to Assumption 2.

Assumption 2: articulation vs. vocalization – the emergence of syllables, vowels and consonants

The simplest sound productions on the basis of the air-stream mechanisms, especially the pulmonic one, set out under Assumption 1, are elementary vocalizations, as they are found in other species, particularly in primates, as well. But human language entails more: it is articulated speech, i.e. short-term local gestures are interspersed into more long-term global movements, articulation rides on vocalization. The precursor to this is the well-known lipsmack, which is a widespread communicative gesture in higher primates, consisting of mandibular opening-closing movements on an oral or a pulmonic mechanism, combined with phonation in 'girney'. It constit-utes a direct link back to vegetative suction and forward to language clicks and labio-velars.

In hominids this basic pattern has been expanded in the syntagmatic as well as in the paradigmatic plane

- by the sequencing of slow mandibular, linguodorsal, labial movement complexes
- by varying these long-time motion units
  - in the vertical dimension: different degrees of mandibular and linguodorsal elevation
  - in the horizontal dimension: front to back linguodorsal and labial movements
- by varying the short-time action inside the global dynamics
  - through extreme strictures on the three air-streams: closure (complete, or central with lateral opening), turbulence narrowing (central/lateral)
  - through velic raising/lowering.

This elaboration of an elementary mandibular opening-closing has resulted in the emergence of concatenative syllables, i.e. motion units of long time constants with short-time-constant articulatory separators, and vowels/consonants inside these units, representing the slow and the faster gestures, respectively. In this connection MacNeilage talks about segmental 'content' in syllable 'frames' [13]. The basic consonant types emerging from this expansion of the fundamental dynamic articulation unit are stops, fricatives, nasals and liquids. The ideal syllable contains a complete articulatory contrast: extreme mandibular opening as well as linguodorsal lowering for the slow setting, and complete supraglottal closure, combined with velic raising, for the faster gesture, resulting in an interruption of the air-stream. This interruption can also be brought about more quickly and efficiently by a glottal closure, which is the most elementary means of cutting off the pulmonic air-stream, being also rooted in vegetative function.

In this connection we must also mention the use of assertion and negation 'particles' instead of the words for 'yes' and 'no', found in many languages. Some of these versions can be transcribed in IPA notation as [a'ha], [a'ɦa], [m'm̥m], [m'm̤m], [ã'hã], [ã'ɦã] (for assertion), ['aʔa], ['mʔm], [' ãʔã] (for negation). They follow the same principle of elementary sound production as the hesitation 'particles' discussed under Assumption 1: an oral or a nasal air-stream or a superposition of both is the carrier for non-lexical vocalizations with situational meaning, and again the degree and shape of mouth opening are not crucial, as long as there is a central or open vowel quality, which once more does not coincide with the vowel phonemes of the particular language.

What is, however, different in these 'particles' from the shaping of hesitations is the introduction of a syllable structure, which represents a further step in the evolution of human language - from vocalization to articulation: these sound productions are thus more speech-like. In both cases the chunking of the air flow into syllabic units is accomplished by glottal activity. The *gradual division*, achieved by a change from modal to breathy voice or to voicelessness and back to modal voice is linked with the meaning *'yes'*, the *abrupt division* by a glottal stop with the meaning *'no'*. In German the assertion is additionally associated with a medial pitch peak or a pitch valley on the second syllable, the negation with an early pitch peak on the first [14]. These different links make a high or a low pitch on these 'particles' perceptually salient. According to an ethological perspective [15,16] high/low pitches have become associated in the course of evolution across species with small/large body size of the vocalizers and secondarily with their submissive/threatening behaviour. A 'frequency code' has thus developed for the signalling of the attitudes of submission and dominance by high and low pitch, respectively. The elementary syllabification of a pulmonic air-stream by basic glottal mechanisms can thus be supplemented by an equally fundamental use of pitch to strengthen the meanings of acceptance and refusal in these non-lexical, but articulated speech interactions.

Assumption 3: coronals – typical consonants

Tongue tip/blade articulations (coronals: apicals/(sub)laminals) are ideal riders on syllable frames because they can be executed faster than tongue dorsum and lip movements. Their production is governed by the short and fast intrinsic tongue muscles as well as the small movable mass, as against the extrinsic tongue or the lip muscles responsible for moving the greater mass of the tongue body or lips. Coronals thus have a shorter time constant than labials or dorsals, and can therefore function as optimal syllable separators because their interaction with the slow global dorsal and lip gestures can be kept minimal, and the articulatory contrast between the two movement types can thus be maximized. Consequently consonants may be assumed to be primarily coronal, vowels non-coronal (dorsal). This is indeed supported by the general occurrence of coronal consonants in the languages of the world, especially in the fricative, nasal and liquid categories, where a unique class element is always coronal. And it is further supported by the scarcity of coronal (apical, laminal or sublaminal/retroflex) vowels.

The frequently observed occurrence of [t] instead of [k] in the speech development of children may also have an articulatory, vegetative base, in addition to its possible origin in a reduced acoustic/auditory contrast between the two sounds. [t] and [k] may be looked upon as phylo-genetic and ontogenetic developments from the [t‿k] double articulation needed in sucking, which is an established reflex right from birth. This would account for the interchangeability of the coronal and dorsal stop articulations in speech sound development. The preference of the coronal element could in turn find an explanation in the greater afferent control of the tongue tip area as against the tongue body, because of the richer supply of kinesthetic and tactile sensors.

Assumption 4: constraints on phonation – obstruents and sonorants

In stop and turbulent fricative strictures vocal fold vibration can only be maintained if the subglottal-supraglottal pressure differential is actively controlled through, e.g., cavity expansion (for example larynx lowering). In the case of voiced fricatives this is particularly costly because maintaining friction simultaneously with glottal pulsing requires the generation of a strong air-stream, which in turn presupposes glottal abduction and raises the supraglottal pressure, both processes being opposed to voicing. From the point of view of production economy stops and fricatives are therefore naturally unvoiced; they become articulatorily more elaborated if they are voiced. In historical sound change, on the other hand, voiced stops tend to be simplified in terms of this concept of articulatory ease, i.e. they become unvoiced. Voiced fricatives may be simplified in two ways: they either lose their voicing, or their friction, becoming approximants. Both processes are widespread in the languages of the world.

In the case of non-stop and non-fricative consonants voicing is not impeded, and they are generally voiced, thus producing a maximal articulatory contrast with the other consonant class. In the case of nasals, lack of vocal fold vibration also minimizes the perceptual distinctiveness, and voiceless nasals are, therefore, not basic consonant types.

This difference in the control of phonation in different consonant productions, due to biomechanical constraints, leads to the emergence of the consonant classes of obstruents and sonorants.

Assumption 5: a force contrast in obstruents – fortis/lenis

A force contrast is associated with the formation and release of pulmonic obstruents. In the case of plosive articulations it manifests itself as a fast/slow occlusion, a long/short hold and a strong/weak release. In fricatives the movement into the turbulence stricture is again fast/slow, and the turbulence itself is characterized by high/low air-stream energy. This leads to the fortis and lenis categories. In stops either the one or the other may be intensified by subsidiary phonation components to generate more elaborated consonant productions:

– the fortis class may be
  -- aspirated or
  -- glottalized or
  -- produced with a glottalic air-stream mechanism, intensifying the stop category still further by heightening glottalization through an upward larynx movement;
– contrariwise the lenis class may be
  -- fully voiced or
  -- breathy voiced or
  -- fully voiced combined with an ingressive glottalic air-stream.

These articulatory intensifications at the same time increase perceptual distinctivity.

Assumption 6: degrees of opening vs. front-back in vowels

Degrees of mandible and linguodorsal raising/lowering dominate over the front-back dimension, because along the latter the tongue does not have the synergy of the mandible, and the gesture is therefore more costly with regard to biomechanics and neuromuscular control. So from the point of articulatory economy, vowel differentiation is to be expected to be greater in the vertical than in the horizontal parameter, which is supported by the empirical data from the world's languages. The principle of sufficient perceptual contrast then operates on the basis of these production constraints. This two-stage evolution model, in which articulatory factors determine the sound space for auditory selection will generate vowel systems in greater conformity with empirical findings than Lindblom's more exclusive focus on perceptual determinants.

Assumption 7: the labial and lingual parameters in consonants

Assumption 3 has underlined the basic status of coronals in consonant systems. However, cutting off the air-stream is an elementary gesture of any one of the active articulators, i.e. the lips, the tongue tip/blade and the tongue body (over and above the glottis, cf. Assumption 2), or of both lips and tongue body. The latter, double articulation provides a link with clicks (see Assumption 1). The same considerations apply to the formation of an oral closure combined with velic opening, i. e. to nasals, at these articulators.

In the case of plosives the perceptual salience is relatively weak, if not supported by additional phonatory features, and it does not seem to differ much between any two of them, as is evidenced by confusions in noise [17]. In the case of nasals the auditory differentiation between the three places is much weaker, and in particular the labial and dorsal articulations are poorly contrasted. The empirical data of the world's languages show the preponderance of three-term plosive and of two-term nasal systems, including [ptk]/[bdg] and [mn], respectively, thus supporting the general assumptions about articulatory and perceptual constraints.

In fricatives one would expect the same distribution across the articulators lips, tongue tip, tongue body (and glottis), because, apart from the weighting of coronals, there does not seem to be any further a priori reason rooted in production for preferring certain turbulence strictures. But here the perceptual salience associated with each one varies very unevenly from weakest for [f] to strongest for [s] and [ʃ]; i. e. the coronals, which are already given articulatory weight in consonant production (see Assumption 3), are also, at least partly, singled out on an auditory basis. This can explain the general occurrence of [s] in the fricative systems of the world's languages, either as the only element or besides others, especially [f] and [ʃ], with dorsal fricatives being quite rare.

Just as [ʔ] constitutes a very basic stoppage of the pulmonic air-stream, [h] provides a similarly elementary turbulence stricture at the glottis. Its occurrence in the majority of languages testifies to this. However, its by no means rare absence from consonant systems requires a separate explanation. [h], in contradistinction from [ʔ], is not an optimal syllable separator because, due to the lack of localised supraglottal articulation on an uninterrupted air-stream, it is integrated into the slow syllable dynamics. Since its differentiation from the context is by glottal abduction only this parameter may become adjusted as well, resulting in breathy voice, which is perceptually weak and may thus disappear. This is a case of articulatory reduction of short-time-constant consonants and their submergence in long-time-constant syllable frames (see Assumption 10).

As regards laterals their labial production generates sound that is not substantially different from a central fricative or approximant. Auditory selection thus eliminates it. This only leaves coronal and dorsal types. As argued in Assumption 3 the coronal articulation is rated higher, and according to Assumption 9 the dorsal one is to be treated as an elaboration, partly originating from coarticulation in syllable frames.

Finally, trills can be generated most efficiently by the tongue tip: the lips have greater mass and the uvula less muscular control for quick movement. This explains the almost universal occurrence of coronal productions.

Assumption 8: the basic vowel and consonant systems

As a conclusion from the foregoing discussion we can postulate the following
BASIC SPEECH SOUND SYSTEMS:

VOWELS			
high vowels	i		u
low vowels		a	

CONSONANTS				
clicks	ʘ	ǂ		
plosives	p	t	k	ʔ
	b	d	g	
fricatives	s	h		
nasals	m	n		
laterals	l			
trills	r			

This gives a frequency of roughly 70% to 30% for obstruents vs. sonorants. As Lindblom and Maddieson [7] have pointed out the elaboration of these two basic consonant system parts occurs proportionally so that the basic relation stays the same across all the systems found in the world's languages.

Assumption 9: elaboration of the basic systems

The basic vowel system is elaborated in the vertical dimension in the first instance, leading successively to five and seven-term vowel systems, the former being the most common in languages; in the second instance there are also elaborations along the horizontal parameter. In both cases the initial articulatory weighting of the vertical dimension is further controlled by the principle of sufficient auditory contrast.

The basic consonant system is elaborated in two ways:

(1) There is further differentiation in the horizontal plane
   - separating bilabial from labiodental and dental from (post)alveolar/retroflex articulations
   - adding the labial and dorsal tongue gestures in fricatives

(2) More elaborations originate from coarticulation, i. e. from integration of 'content' into 'frame'
   - palatal fricatives, nasals, laterals for dorsals in front vowel contexts
   - uvular fricatives for dorsals in low back vowel contexts
   - secondary articulations of labialization, palatalization, velarization and pharyngealization in the appropriate vowel contexts
   - creaky voice phonation (glottalization) in consonants in the context of plosive-related glottal stops.

The coarticulatory origin within the syllable also applies to vowels:

- nasal vowels always develop in nasal consonant contexts
- retroflex vowels originate in the neighbourhood of retroflex consonants
- creaky vowels in glottal stop onsets
- breathy vowels in breathy consonant contexts.

Assumption 10: articulatory reduction

The initial separation of 'content' from 'frame' is counteracted by articulatory reduction, integrating the former into the latter for economy of effort:

- intervocalic lenis stop syllable separators become approximants
- glottal stops are reduced to low-frequency irregular glottal pulsing superimposed on vowels and sonorants
- for nasal consonants the oral occlusion is no longer effected, resulting in nasal vowels
- coronal stops and nasals are assimilated to the labial or the dorsal articulator, thus removing the consonantal specificity and adjusting to the always ongoing lip and tongue movements
- intervocalic lenis fricatives become approximants
- trills are reduced to taps and further to approximants, which may then be integrated into the vowels.

This is just a small selection of possible processes, which may be multiplied and abound in spontaneous speech. But in all these cases the extent of integration of short-time into long-time movements is checked through the demands for perceptual contrast imposed by the listener in the communicative situation. These synchronic reduction processes may then lead to petrified patterns in historical sound change, when the perceptual check by listeners is relaxed.

The integration of 'content' into 'frame' may also result in new content elements, such as palatals, nasal vowels and glottalized sonorants. So what we find in languages at any particular point in time is not only the result of long-term phylogenetic evolution of human sound systems, but also of much more short-term historical changes due to the striking of a new balance between articulatory ease and perceptual distinctivity under specific cultural settings for communication.

Conclusion

The original emergence of 'frame' and 'content', separating two time constants in articulatory units, was the principal step in the evolution of human language. But the boundary between the two categories can become blurred in listener-tolerated reduction of articulatory effort. This may in turn lead to the emergence of new 'content' elements, and the cyclic change from 'content' to 'frame' to 'content' may continually repeat itself in historical time, under changes in the complementary control of articulatory effort and perceptual contrast.

This means that consonant and vowel system surveys, like [11], of the world's languages can only be a rough guide for questions of human language evolution and of language universals, because the data represent different historical stages in the 'content'-'frame' cycles, which are bound to cover up general evolutionary trends of human language, if taken at face value, clicks and glottalized consonants being cases in point.

For example, if German were an unwritten language a field-worker, coming across the word pair 'können' [kœnn] and 'könnten' [kœn̰n] [18,19], would conclude that German has a consonant opposition between clear and creaky-voice nasals, which comes under the subsidiary, more elaborated features of human sound systems and is relegated to a more peripheral occurrence. But what we are really dealing with here is the most elementary cutting off of the pulmonic air-stream by glottal closure, which may in turn be weakened to irregular pulsing, i.e. creaky voice, thus creating a secondary contrast between two types of phonation. What is important from the evolutionary point of view is the basic glottal action, not the 'elaborated' sound contrast.

Considering the procedure adopted for collecting the UPSID data base, examples like this one from German may be expected to be quite numerous, and therefore valid divisions into evolutionary basic, elaborated and complex sets cannot be made on the strength of this corpus. Moreover, the data focus entirely on the paradigmatic aspect of sound systems, ignoring the syntagmatic domain completely, thus missing the essential distinction between 'frame' and 'content'. Finally, the approach is exclusively linear, segmental, phonemic and thus also excludes componential aspects of sound systems, which are associated with the 'frame'.

So future research into language evolution, universals and typology will have to put greater focus on syntagmatic aspects of speech production. In particular, paradigmatic systems need to be supplemented by sound patterns of connected, especially spontaneous speech. And this expanded range of phenomena needs more adequate phonological tools that will give equal importance to componential elements beside segmental ones in a 'complementary phonology' [20].

References

[1] Liljencrants, J. & Lindblom, B. (1972): Numeric simulation of vowel quality systems: the role of perceptual contrast. *Language* 48: 839-862

[2] Lindblom, B. (1975): Experiments in sound structure. Paper read at the 8th International Congress of Phonetic Sciences in Leeds

[3] Lindblom, B. (1984): Can the models of evolutionary biology be applied to phonetic problems? In: *Proc. Xth ICPhS Utrecht* (M. P. R. van den Broecke & A. Cohen, eds.), Foris: Dordrecht, 67-81

[4] Lindblom, B. (1986): Phonetic universals in vowel systems. In: *Experimental Phonology* (J. J. Ohala & J. J. Jaeger, eds.), Academic Press: Orlando, 13-44

[5] Lindblom, B. (1990a): Explaining phonetic variation: a sketch of the H and H theory. In: *Speech Production and Speech Modelling* (W. J. Hardcastle & A. Marchal, eds.), Kluwer: Dordrecht, 403-439

[6] Lindblom, B. (1990b): On the notion of 'possible speech sound'. *Journal of Phonetics* 18: 135-152

[7] Lindblom, B. & Maddieson, I. (1988): Phonetic universals in consonant systems. In: *Language, Speech and Mind* (L. M. Hyman & C. N. Li, eds.), Routledge: London, 62-78

[8] Newman, F. R. (1980): *Mouthsounds*. Workman Publishing: New York

[9] Classe, A. (1957): Phonetics of the silbo Gomero. *Archivum Linguisticum* 9, 44-61

[10] Pike, K. L. (1943): *Phonetics*. University of Michigan Press: Ann Arbor

[11] Maddieson, I. (1984): *Patterns of Sounds.* CUP: Cambridge

[12] Pätzold, M., and Simpson, A.: An acoustic analysis of hesitation particles in German. *Proc. ICPhS Stockholm*, Vol. 3, 512-515

[13] MacNeilage, P. F., and Davis, B.L. (1990): Acquisition of speech production: the achievement of segmental independence. In: *Speech Production and Speech Modelling* (W. J. Hardcastle & A. Marchal, eds.), Kluwer: Dordrecht, 55-68

[14] Kohler, K.J. (1991): Prosody in speech synthesis: the interplay between basic research and TTS application. *J. of Phonetics* 19, 121-138

[15] Ohala, J. (1983): Cross-language use of pitch: an ethological view. *Phonetica* 40: 1-18

[16] Ohala, J. (1984): An ethological perspective on common cross-language utilization of F0 of voice. *Phonetica* 41: 1-16

[17] Miller, G. A., & Nicely, F. E. (1955): An analysis of perceptual confusions among some English consonants. *J. acoust. Soc. Am.* 27: 338-352

[18] Kohler, K.J. (1994), Glottal stops and glottalization in German. Data and theory of connected speech processes. *Phonetica* 51: 38-51

[19] Kohler, K.J. (1995): The realization of plosives in nasel/lateral environments in spontaneous speech in German. *Proc. XIIIth ICPhS Stockholm*, Vol. 2, 210-213

[20] Kohler, K.J. (1994): Complementary phonology: a theoretical frame for labelling an acoustic data base of dialogues. *Proc. ICSLP94,* vol. 1, 427-430