Symposium for

# Bill Barry's 60th anniversary

with contributions from his teachers, students and co-workers

Institute of Phonetics,
Saarland University, Saarbrücken

25th April 2003

# Programme

| | |
|---|---|
| 9.00 - 10.00 | Coffee |
| 10.00 - 10.30 | Francis Nolan, Cambridge:<br>*Linguistic and non-linguistic determinants of pitch range* |
| 10.30 - 11.30 | Klaus Kohler, Kiel:<br>*Prosody revisited – "function", "time", and the "listener" in intonational phonology* |
| 11.30 - 12.00 | Coffee with Demonstration<br>Marc Schröder, Saarbrücken:<br>*The German text-to-speech synthesis system MARY* |
| 12.00 - 12.30 | Paul Dalsgaard, Aalborg:<br>*On acoustic-phonetic features and their use in multi-lingual speech technology research* |
| 12.30 - 13.00 | Jacques Koreman & Manfred Pützer, Saarbrücken:<br>*The usability of perceptual ratings of voice quality* |
| 13.00 - 14.00 | Lunch |
| 14.00 - 14.30 | Dafydd Gibbon, Bielefeld:<br>*Computational phonetics: empirical comparison of prosodic and syntax tree* |
| 14.30 - 15.00 | Henning Reetz, Konstanz:<br>*It's the model, not the data* |
| 15.00 - 15.30 | Bistra Andreeva & Jacques Koreman, Saarbrücken:<br>*Phonatory demarcations of intonation phrases in Bulgarian* |

| | |
|---|---|
| 15.30 - 16.00 | Tea with Demonstration<br>Ralf Benzmüller, Bochum:<br>*Logox- a flexible and compact text-to-speech system* |
| 16.00 - 16.30 | Hartwig Eckert, Flensburg:<br>*Phonetic transcription in TEFL* |
| 16.30 - 17.00 | Michela Russo, Paris & Saarbrücken:<br>*Chute du schwa final: interaction entre structure segmentale et caractère rythmique* |
| 17.00 - 17.30 | Jürgen Trouvain, Saarbrücken:<br>*Is phonetics a laughing matter?* |
| *Chair* | Martine Grice, Cologne & Saarbrücken |

All talks will be given in the Seminar room (ground floor),
Building 17.2 , Saarland University

| | |
|---|---|
| 19.00 | Dinner at Gasthaus "Zum Stiefel", Saarbrücken, St. Johanner Markt |

# LINGUISTIC AND NON-LINGUISTIC DETERMINANTS OF PITCH RANGE

Francis Nolan

Department of Linguistics, University of Cambridge

fjn1@cus.cam.ac.uk

This talk will draw together a number of strands of research in Cambridge related to pitch range. ‘Pitch range’ is taken to cover a number of distinct but related phenomena, namely pitch level (the placing of an utterance within a speaker’s tessitura or physiologically determined comfortable pitch range), pitch span (the general pitch extent of an utterance), and pitch excursion (the pitch extent of a more localised event within an utterance). Pitch range is challenging because it rather awkwardly straddles the linguistic and the non-linguistic.

First, I will deal with the effects on pitch range of ‘speaking up’ and ‘local emphasis’. Ladd and Terken (1995) showed for Dutch that whereas local emphasis raised only the pitch peak associated with an accented syllable, ‘speaking up’ against noise raised all parts of the pitch contour, with the near exception of the ‘final low’. I will report a partial replication of this work by Jan Henning Schultze (2003) for Northern Irish English, which is characterised by low (nuclear) accents and a final high plateau. The questions are (a) whether these modifications cause the L* to go down or up, and (b) whether there is an ‘anchor’ comparable to the final low.

I will then show how using noise to manipulate a speaker’s pitch range can resolve disputes about whether pitch accents are present or not in certain intonational phrases. The illustration comes from work by Lluisa Astruc (In Progress) on Catalan, including reported speech (‘… , she responded’) and right dislocations (‘[He] studies mathematics, my brother-in-law’). The idea is that underlying pitch accents lurking in reduced-excursion parts of the utterance might be ‘flushed out’ when the pitch range is expanded. I will also refer to an experiment on intonational equivalence (Nolan, 2003) in which speakers imitated utterances in three different pitch spans. Knight (2002), using data from that experiment, showed that the ‘plateau’ often associated with an English nuclear syllable is carefully controlled as pitch span varies.

Finally, I will deal with the interaction of pitch range with grammar. Chia-Hui Sung (2002) showed that different kinds of question in Taiwanese are marked by different degrees of raising of pitch level, uptilt of ‘top-line’, and expansion of early and late excursions. The interaction with morpho-syntactic marking (absent in ‘declarative’ questions, and final or medial in different types of yes-no and wh-questions) suggests not only a trading relationship between pitch range and morpho-syntactic marking, but also a more specific hypothesis that intonational cues are oriented towards the listener’s real-time processing needs. This hypothesis has been tested for a non tone-language by Eva Liina Asu (In Progress).

Asu, E.L. (In Progress). *The intonation of Estonian*. PhD dissertation, University of Cambridge.

Astruc, L. (In Progress). *The intonation of sentence-external elements*. PhD dissertation, University of Cambridge.

Knight, R-A. (2002). The effect of pitch span on intonational plateaux. Proc. *Speech Prosody 2002*, pp. 439-442.

Ladd, D.R. & Terken, J. (1995). Modelling intra- and inter-speaker pitch range variation. Proc. *13th Int. Congress of Phonetic Sciences*, pp. 386-389.

Nolan, F. (2003). Intonational equivalence: an experimental evaluation of pitch scales. To appear in Proc. *15th Int. Congress of Phonetic Sciences*.

Schulze, J.H. (2003). *Pitch range variation in Northern Irish English due to ‘speaking up’ and local emphasis*. Long essay for MPhil in Linguistics, University of Cambridge.

# PROSODY REVISITED
# "FUNCTION", "TIME", AND THE "LISTENER" IN INTONATIONAL PHONOLOGY

Klaus J. Kohler

Instutute of Phonetics and Digital Signal Processing (IPDS), University of Kiel

kjk@ipds.uni-kiel.de

For his habilitation lecture and colloquium in 1980, Bill Barry chose the subject "Prosodic functions revisited again!", which was subsequently published in Phonetica. His paper marks the beginning of a prosodic research focus at IPDS Kiel on the phonetic form - meaning relationship of intonation, which tries to encompass all types of functions that are relevant in speech communication. Bill Barry stresses two points:

(a) to consider function systematically in the very analysis of prosody (as against the often anecdotal reference in the British School or even its exclusion in such phonological models as the ones by Halliday, the Dutch School or Pierrehumbert),

(b) to go beyond the traditionally overemphasized linguistic function (sentence modality, focus, syntactic phrasing) and to include speaker attitude, the interaction function (phrasing between speakers in dialogue) and the guide function (the importance of prosodic continuity and prosodic structure of the preceding context for perceiving messages in real speech situations).

The recognition of different functions and the expansion of their scope, later became the prerequisite for the prosodic analysis of spontaneous speech, where the whole spectrum of communicative features of meaning interact and produce prosodic exponents that would be forced into a Procrustean bed, or cannot be handled adequately at all, by prosodic categories established entirely within linguistic structures. The guide function is, for example, operative in production across false starts or hesitations and subsequent pauses where the global intonation pattern continues after the interruptions as if they had not occurred. In turn, the guide function brings in the listener as an essential category in the modelling of prosody. As phonology sets up linguistic units independently of a speaker and a listener, prosodic categories within a strictly linguistic framework lack essential aspects of the communicative situation. But the listener is also important at the metalinguistic level of data processing (e.g. labelling), where modern lab phonology approaches tend to ignore the ear and to rely on measuring acoustic parameters, which are then coded, e.g., in binary terms H/L, as sequences of local points.

In the winter semester of 1982/3, Bill and I gave an advanced seminar series on "Functions of intonation and their phonetic manifestations". In these classes, the functional categories of semantic focus ('opening' vs. 'concluding argument', and 'assessing the situation in contrast to one's own expectation') as well as of the listen-er were further elaborated in the first perception experiments with the German sentence "Sie hat ja gelogen." A new experimental paradigm was initiated by shifting a whole F0 peak contour in equal time steps through a seg-mentally constant utterance, and by testing the resulting perceptual pitch changes in the stimulus series as well as their associated semantic features. This paradigm introduced two further categories into intonation research:

(1) synchronization of phonatory (pitch) and vocal tract (articulatory) time courses; thus time became a feature of intonational phonology (just as, several years later, articulatory phonology incorporated the temporal aspect into the segmental strand), (2) the essential intonational components were assumed to be more global contours (peaks, valleys), rather than local pitch levels.

The paradigm was later elaborated in a long-term German Research Council project "Form and function of intonation peaks" within the focus programme "Forms and functions of intonation". The perception results established three peak positions for the prosodic phonology of German, with a categorical change from early to medial peak, and a more gradual one to the late peak. The semantic changes paraphrased above are also clear-cut. On this basis the Kiel Intonation Model (KIM) was developed.

Latterly, a second time dimension was introduced into the experimental paradigm in two MA dissertations: the internal timing (shape) of the peak contour, i.e. the speeds of the peak rise and the peak fall (each slow vs. fast). Oliver Niebuhr tested the effects of this category with extensive listening experiments in German and found strong interactions with the position change from an early to a medial peak: a fast rise + slow fall produces a significantly earlier percept change from early to medial, compared with a slow rise + fast fall. Similarly, Tamara Khromovskikh tested the internal timing effects on the coding of statements vs. yes-no questions (which lack syntactic markers) in Russian. Here the two sentence categories are coded by early vs. late peak positions, combined with slow rise + fast fall vs. fast rise + slow fall (and by additional lower vs. higher peak value).

These data will be further discussed and supplemented by data from Bulgarian (narrow-focus statement vs. question: Bistra Andreeva) and Italian (wide vs. narrow focus in Pisa Italian: Barbara Gili Fivela; statement vs. question: Mariapaola d'Imperio). Finally, an explanation will be presented as to why the perceptual changes may depend on the synchronization and the internal timing in the way they do, and why specific feature combinations in peaks may be used to differentiate questions from statements. This will be done with reference to two theoretical considerations: auditory contrast in contours at specific syllable points (auditory enhancement), and the frequency code. The former may also explain why valley shifts from early to late positions so far have not replicated the results found in peak shifts.

# THE GERMAN TEXT-TO-SPEECH SYNTHESIS SYSTEM MARY

Marc Schröder

German Research Centre for Artificial Intelligence (DFKI), Saarbrücken

and Institute of Phonetics, Saarland University, Saarbrücken

schroed@dfki.de

The German text-to-speech synthesis system MARY is demonstrated. The system's main features, namely a modular design and an XML-based system-internal data representation, are pointed out. An interface allowing the user to access and modify intermediate processing steps without the need for a technical understanding of the system is shown, along with examples of how this interface can be put to use in research, development and teaching. The usefulness of the modular and transparent design approach is further illustrated with an interface for emotional speech synthesis.

Schröder, M. & Trouvain, J. (to appear). The German Text-to-Speech Synthesis System MARY: A Tool for Research, Development and Teaching. *International Journal of Speech Technology*, Special Issue following the 4th ISCA Workshop on Speech Synthesis.

URL: http://mary.dfki.de

# ON ACOUSTIC-PHONETIC FEATURES AND THEIR USE IN MULTI-LINGUAL SPEECH TECHNOLOGY RESEARCH

Paul Dalsgaard

Center for PersonKommunikation (CPK), Aalborg University

pd@cpk.auc.dk

This presentation describes a discussion framework addressing the general need to bridge the gap of understanding between theoretical concepts as given in traditional Phonetics and Phonology and their practical implementation and utilisation within Speech Technology - as seen through the eyes of a speech technologist. The question of linguistic equivalence within acoustic variability is pursued to the logical extreme of multi-lingual processing, and the viability of cross-language feature definition in ASR is examined. The aims of the presentation are both theoretical and practical: theoretical in that a brief discussion is offered of the phonetic and phonological issues that implicitly underlie the reported experiments; practical in that it presents the results of a multi-lingual database experiment aimed at analysing the potential 'power' of merging acoustically similar sounds across a number of languages in a language identification experiment.

Three different levels of approach may be usefully distinguished. These are, firstly, the implicit application of phonetic and/linguistic knowledge in the definition of the speech-material used in training, secondly, the implicit incorporation of phonetic knowledge in the development of signal-processing and training methods, and thirdly, the explicit inclusion of phonetic knowledge as the (joint) basis of stochastic processing.

*At the first level*, the recognition of major allophonic variants and their separate treatment as training units implies an acceptance of the fundamentally abstract nature of the phoneme. Similarly, the use of diphone- and triphone-based training (i.e. training units consisting of separate two- and three-segment strings) acknowledges the inherently dynamic nature of the acoustic signal resulting from articulatory movements from one sound to another.

*At the second level*, it is probably fair to say that all the stochastic methods cater for some aspects of phonetic variability, but they are all subject to the constraints imposed by the selection of training and recognition units. HMM modelling allows for the same sort of temporal variation, catering also for missing elements (states), which can theoretically deal with cases of structural modification. Neural net approaches offered the great advantage of "learning" from the actual input, but in fact they are not inherently more able to deal with the variation within and overlap between defined recognition units than other approaches.

*At level three*, no general trend can be determined because the examples are all the product of individuals' insights, motives and preferences.

*Poly- and Monophonemes*

Behind the discussion of acoustic-phonetic variation in the manifestation of features and phonemes lies a doubled-edged question: 1) does the feature retain enough phonetic identity *within* a language to be practically viable; 2) is the variation in the manifestation of a feature *across* languages restricted enough to allow it also to function multi-lingually. Within our concern for speech-material economy, the definition of cross-language equivalence is crucial.

Two strategies have been examined so far for defining such equivalences, the first of which was based on the use of phonemic distinctions using the feature specific-ations, the second on a purely data-driven approach using the acoustic speech material only.

*Phonemic distinction strategy*

Using the first strategy, phonetically based decisions were required at two levels. The first level precedes the phonetic-feature definition of the sounds in the individual languages; in view of acoustic differences between positional variants of any one phoneme, a considered choice of the most representative variant is required. Thus, although in Danish, English and German, sonorant consonants such as the nasals and the /l/ and /r/ may be voiceless for a considerable part of their temporal extension in stressed syllable-initial clusters after voiceless sibilants and/or plosives, it can be assu-med that the occurrence of these variants is less frequent than the sum of other variants which are not devoiced, and they can be allocated the [+sonorant] feature as a *phonetic* feature (not just the *phonological class* feature) characterising the acoustic property of voicing.

*Data-driven strategy*

The second strategy is based on discriminative cues in the acoustic signal focussing on the spectral and temporal structure of the individual sounds. These characteristics are modelled using three-state CDHMMs described by a set of cepstral parameters. After training the models for all speech sounds in the combined speech corpus, a test session is conducted in which the confusions between the speech sounds are quantified.

# THE USABILITY OF PERCEPTUAL RATINGS OF VOICE QUALITY

Jacques Koreman & Manfred Pützer

Institute of Phonetics, Saarland University, Saarbrücken, Germany

{koreman,puetzer}@coli.uni-sb.de

In this study, roughness, breathiness and hoarseness scores (RBH) of eight raters are compared for continuous vowels [i:,a:,u:]. The vowels are spoken by one 'normal' speaker group and two patient groups with an adduction deficiency due to a unilateral vocal fold paralysis. Interrater correlations are computed for the RBH scores, and then distinctions between the three speaker groups are evaluated on the basis of the scores. The results show that although interrater agreement for the perceptual scores is only moderate, similar distinctions between the three speaker groups are made by each rater and therefore support the applicability of perceptual rating systems like the RBH system in clinical practice.

Dejonckere, P.H., Remacle, M., Fresnel-Elbaz, E., Woisard, V., Crevier-Buchman, L. & Millet B. (1996). Differentiated perceptual evaluation of pathological voice quality: reliability and correlations with acoustic measurements. *Revue de Laryngologie Otologie Rhinologie* 117(3), pp. 219-224.

Gerratt, B.R. & Kreiman, J. (2000). Theoretical and methodological development in the study of pathological voice quality. *Journal of Phonetics* 28, pp. 335-342.

Hanson, H.M. & Chuang, E.S. (1999). Glottal characteristics of male speakers: acoustic correlates and comparison with female data. *Journal of the Acoustical Society of America* 106(2), pp.1064-1077.

Koreman, J. & Pützer, M. (1997). Finding correlates of vocal fold adduction deficiencies. Working Papers Institut für Phonetik Saarbrücken *Phonus* 3, pp. 155-178.

Kreiman, J. & Gerratt, B.R. (2000). Sources of listener disagreement in voice quality assessment. *Journal of the Acoustical Society of America* 108(4), pp. 1867-1876.

Nawka, T. & Anders, L.Ch. (1996). *Die auditive Bewertung heiserer Stimmen nach dem RBH-System* (Audio-CDs mit Stimmbeispielen). Thieme Verlag, Stuttgart.

Nawka, T., Anders C. & Wendler, J. (1994). Die auditive Beurteilung heiserer Stimmen nach dem RBH-System. *Sprache Stimme Gehör* 18, pp. 130-133.

Pützer, M. & Koreman, J. (1997). A German database of patterns of pathological vocal fold vibration. Working Papers Institut für Phonetik Saarbrücken *Phonus* 3, pp. 143-153.

Sataloff, R. (1997). *Professional Voice. The Science and Art of Clinical Care*. Singular Publishing Group, San Diego.

Wendler, J., Seidner, W., Kittel, G. & Eysholdt, U. (1996). *Lehrbuch der Phoniatrie und Pädaudiologie*. Thieme Verlag, Stuttgart.

Wirth, G. (1995). *Stimmstörungen*. Deutscher Ärzteverlag, Köln.

Wolfe, V., Fitch, J. & David, M. (1997). Acoustic measures of dysphonic severity across and within voice types. *Folia Phoniatrica et Logopaedica* 49, pp. 292-299.

# COMPUTATIONAL PHONETICS: EMPIRICAL COMPARISON OF PROSODIC AND SYNTAX TREES

Dafydd Gibbon

Faculty of Linguistics and Literary Studies, University of Bielefeld

gibbon@spectrum.uni-bielefeld.de

Empirical study of the syntax-prosody relation is hampered by the fact that current timing models (including rhythm models) are essentially linear, while textual structure is hierarchical. The present contribution describes a syntax-prosody comparison heuristic based on two new algorithms: Time Tree Induction, TTI, for building a prosodic treebank from time-annotated speech data, and Tree Similarity Indexing, TSI, for comparing syntactic trees with the prosodic trees. In a preliminary study, two parametrisations of the TTI algorithm, for different tree branching conditions, are applied to sentences taken from a read-aloud narrative, and compared with parses of the same sentences, using the TSI. In addition, null-hypotheses in the form of flat bracketing of the sentences are compared. A preference for iambic (heavy rightmost branch) grouping in English, and considerably less clear preferences for a tone language (Ibibio, Lower Cross, Nigeria) are found. The resulting quantitative evidence for syntax-prosody relations has potential applications in speech genre characterisation and in duration models for speech synthesis.

# IT'S THE MODEL, NOT THE DATA

Henning Reetz

Department of Linguistics, University of Konstanz

Henning.Reetz@uni-konstanz.de

Many people in nowadays speech research are entailed in building large databases of speech related data to train and verify statistical speech recognition systems. Despite these are usually Hidden Markov Models, the underlying models are not phonetically or otherwise linguistically motivated but are a stochastic model of events that depend only on their preceding chain of events (and in practice, only on the immediately preceding event). On the other hand, by using a model that inherits information about the nature of speech additional information not available from the raw signal can be added to the system and eventually improve recognition. This line of research has been abandoned nearly completely in the past two decades with the outcome that speech recognition systems have hardly improved despite an tremendous increase of computing power in this time. Because is is hard to predict the future, the talk will focus on the past into an area which had a similar pattern with large databases that became only valuable after a proper model was applied: the astronomy.

# PHONATORY DEMARCARTIONS OF INTONATION PHRASES IN BULGARIAN

Bistra Andreeva & Jacques Koreman

Institute of Phonetics, Saarland University, Saarbrücken

{koreman,andreeva}@coli.uni-sb.de

Vowel devoicing at phrase boundaries is shown to depend on sentence mode (statement, question) and information structure (non-contrastive vs. contrastive and broad vs. narrow focus), which affect the type of nuclear accent and boundary tone.

Analogously to the observations for vowel loss and reductions (Barry & Andreeva, 2000), we never observe complete vowel devoicing with final rising intonation contours (*li*-questions), while it is quite frequent in falling intonation contours (statements and checks).

For the final falling intonation contours (L-L%), the position of the nuclear accent with respect to the intonation phrase boundary is one of the main determinants of phonatory demarcation in the form of vowel devoicing: the further the pitch accent is away from the phrase boundary, the larger the relative devoiced portion of the vowels in the last two syllables in the phrase becomes.

When the nuclear accent is realised on the last content word in the phrase, we found that sentence mode (statement vs. question) and focus type (broad vs. narrow; contrastive vs. non-contrastive) sometimes affect the duration of the relative devoiced vowel portions. But the (lack of) differences could mostly be explained by corresponding (lack of) differences in the phonetic realisation (peak alignment) of phonological categories (accent type).

The focus-associated accent in statements was analysed in Andreeva, Avgustinova & Barry (2001) as a H*, with late peak alignment for contrastive and early peak alignment for non-contrastive utterances. In fact, comparing contrastive and non-contrastive statements with different peak alignments with checks (which also have a late peak alignment) we observe significant differences in vowel devoicing for /i/ and /a/ (*'gramatika')* between statements with early and late peaks, as expected. In the comparison with checks (L*+H), we find that /a/ behaves the same as statements with late peaks, but /i/ does not (less devoicing, as shown by the relative devoiced portions on the next page):

| | | | |
|---|---|---|---|
| /i/: | H* (early) > | H* (late) > | L*+H |
| | 86 | 60 | 31 |
| /a/: | H* (early) > | H* (late), | L*+H |
| | 97 | 43 | 35 |

Inspection of the speech signals and the corresponding F0 tracks shows that in non-final position the H* with a late peak is followed by a fall, so that it may be more appropriate to reanalyse the pitch accent as a bitonal H*+L accent. The reanalysis implies a sharp drop in F0, which enhances devoicing of the following /i/ in 'gramatika'. The L*+H accent in checks not only differs in the steep rise towards the peak, it also has a slower fall, which prevents devoicing of the /i/ (since F0 remains high longer). This does not explain why the relative devoiced portion in /a/ is not also longer in statements with late peaks than in checks. We tentatively hypothesise that the need to audibly signal the fall forbids a strong devoicing of /a/.

In summary, vowel devoicing is a phonatory phrase demarcation phenomenon which depends on the phonetic realisation of nuclear pitch contours.

Andreeva, B., Avgustinova, T. & Barry, W.J. (2001). Link-associated and focus-associated accent patterns in Bulgarian. In: Zybatow, G. Junghanns, U. Mehlhorn, G. & Szucsish, L. (eds) *Current Issues in Formal Slavic Linguistics*. Peter Lang, Frankfurt, pp. 353-364.

Barry, W.J. & Andreeva, B. (2000): Cross-language similarities and differences in spontaneous speech patterns. *Journal of the International Phonetic Association,* 31 (1), pp. 51-66.

# LOGOX
# A FLEXIBLE AND COMPACT TEXT-TO-SPEECH SYSTEM

Ralf Benzmüller

G DATA Software AG, Bochum

Ralf.Benzmueller@gdata.de

Logox is a multilingual text-to-speech system which converts written text into audio signals. This process can be divided into two major complexes: text processing, and speech signal processing.

The first is to provide information about pronunciation and prosodic aspects. In different languages the relevant information differs widely. Therefore a flexible system is needed, which can encode different information structures in language dependent data and rule sets, although the algorithms, that process this information, are the same for all languages. The flexible concept that provides this is referred to as SpeechFont®.

Logox's speech signal processing part is based on the microsegment synthesis (cf. Benzmüller & Barry, 1996) where speech knowledge was applied rather than technical compression algorithms.

*A flexible multilingual system with SpeechFonts*

Modelling the way a person speaks, is not restricted to the voice quality. Articulation, speech rhythm, and intonation are important characteristics of a speaker. This is reflected in the modular architecture of the Logox speech engine. The conversion of text into audio signals is done in a number of steps. Each step is carried out by one module, which takes data from its precedent and after processing, passes them to the subsequent module. Modules are described in table 1.

*Microsegments - a compact sound inventory*

The microsegment synthesis developed so as to produce a recognizable human voice, running on a 386 PC. This implied using a concatenative approach with severe restrictions as to memory requirements and computing time. In order to compress the size of the sound inventory, we followed the rationale to use speech material as often as possible. This implied a variable segmentation. For most consonants, microsegments are phone-sized with only little contextual variation. Plosives and affricates are divided into one closure segment and a variable burst segment. Vowels in C-V-C sequences are segmented as vowel halves, which are used in a generalised way according to 3 main places of articulation (labial, coronal, velar) of the neighbouring consonant. In addition there are "stationary" vowel segments, which enable the use of shorter transitional segments next to /h/, /j/, /?/ and in V-V sequences. This resulted in a compact inventory with 380 microsegments and about 1/6 of the size of a comparable diphone inventory, while preserving the dynamics of the articulatory movements.

| | |
|---|---|
| Interpreter | Text normalization |
| Text FX | Optional: textual replacements |
| Preparser | Conversion of textual formats and numbers |
| Parser | Lexicon with data and rules about pronunciation, phrasing and accentuation |
| Phone FX | Optional: Replacements in the phone string |
| Intonation | Assigning ToBI tones |
| Phone to Mseg | Duration, tone realization and conversion of phones into microsegments |
| Mseg to Audio | Applying duration and f0 values to microsegments |
| Audio FX | Optional: Equalizer, Reverb, Chorus, Flanger, etc. |
| Audio Out | Output to soundcard, file, phone etc. |

Table 1: Modules of a Logox SpeechFont

Benzmüller, R. & Barry, W.J. (1996). Microsegment Synthesis - Economic principles in a low-cost solution. Proc. *Int. Conference on Spoken Language Processing 96 Philadelphia* (4), pp. 2383-2386.

URL: http://www.logox.de

# PHONETIC TRANSCRIPTION IN TEFL

Hartwig Eckert

Department of English, University of Flensburg
eckert@foni.net

All English departments in the German speaking area teach phonetic transcription, and phonetic transcription is used in textbooks for beginners, in some books even for pupils aged eight. It seems to be taken for granted that a knowledge of phonetic symbols and - with more advanced students - the skill of phonetic transcription improves learners' pronunciation. I wish to argue that most teachers and textbooks take a monosystemic view on phonetic transcription in the sense that they do not distinguish between exercises that are extremely useful in some areas and less useful in others and that they use transcriptions that are assumed to be equally helpful to learners all over the world.

I will address the issue of whether students who achieve good results in phonetic transcription tests have a better pronunciation in English, and whether students who do less well in these tests have a strong foreign accent?

I will present results from studies where students' productive and receptive transcription skills were compared to their pronunciation skills and I will argue that phonetic trancription exercises in TEFL should be based (a) on phonetic error analysis of the target group and (b) on empirical data of what exactly we are teaching and testing when we use phonetic transcription in schools and universities.

Brown, A. (ed.) (1992). *Approaches to Pronunciation*. Macmillan, London.

Cunnigham, S. & Bowler, B. (1992). *Headway. Intermediate Pronunciation*. OUP, Oxford.

Celce-Murcia, M. et al. (1996). *Teaching Pronunciation: A Reference for Teachers of English to Speakers of Other Languages*. CUP, Cambridge.

Dalton, C. & Seidlhofer, B. (1994). *Pronunciation*. OUP, Oxford.

Fitzpatrick, F. (1995). *Teacher's guide to practical pronunciation*. Macmillan, London.

Germer, E. (1980). *Didaktik der englischen Aussprache*. Schroedel, Hannover.

Haycraft, B. (1978). *The Teaching of Pronunciation. A classroom guide*. Longman, London.

Hollingsworth, K. & Martin, E. (1994). *Sounds good! Ein unterhaltsames Aussprache-training*. Max Hueber Verlag, Ismaning.

Laroy, C (1995). *Pronunciation. Resourcebook for teachers*. OUP, Oxford.

Pennington, M.C. (1996). *Phonology in English Language Teaching*. Longman, London.

Underhill, A. (1994). *Sound Foundations*. Heinemann, Oxford.

Weiher, E. (1994). *Praktische englische Phonetik. Einführung mit Übungen*. Dümmler, Bonn.

Wells, J.C. (2000): *Pronunciation Dictionary*. 2nd edition. Longman, London.

# CHUTE DU SCHWA FINAL: INTERACTION ENTRE STRUCTURE SEGMENTALE ET CARACTÈRE RYTHMIQUE

Michela Russo

Paris 8/UMR 7023: Structures formelles de langage - Centre Nationale de la Recherche Scientifique (CNRS) and Saarland University, Saarbrücken

miru0003@lycos.com

Avec la désignation d'isochronie accentuelle (IA) et isochronie syllabique (IS), on traduit les sigles de Pike 1945 'stress-timing' et 'syllable-timing'. Le système dichotomique établi par Pike 1945 se base sur l'hypothèse que les langues à IA montrent une tendance à l'isochronie des événements rythmiques. Des études sur l'isochronie n'ont cependant pas confirmé l'existence d'intervalles isochrones dans les langues «isosyllabiques» ou «isoaccentuelles» (cf. Roach, 1982; Dauer, 1983). Les critères structuraux de différentiation de Dauer ainsi que les récentes approches instrumentales (quantitatives) basées sur la logique des critères de Dauer fournissent un cadre pour une différentiation plus fine que la dichotomie traditionnelle isochronie syllabique vs. isochronie acentuelle. Aujourd'hui, il n'est plus question de regularité d'intervalle entre syllabes ou pieds, mais plutôt de variation de durée vocalique et de durée consonantique entre les voyelles (Dauer, 1983). Ces mesures tiennent compte de la structure syllabique dominante d'une langue et de la tendance des voyelles à la réduction.

L'italien est généralement considéré comme une langue isosyllabique, même si certaines études ont déjà apporté des preuves de la réduction des voyelles atones pour cette langue (Farnetani & Kori 1984). Plusieurs dialectes italiens centro-méridionaux présentent une érosion des syllabes atones et tendent pour cette raison vers le pôle isoaccentuel. Cette différence de classification entre l'italien et ses dialectes est plausible au vu de certaines différences structurales: structures syllabiques relativement simple vs. structures syllabiques complexes; absence de réduction des voyelles au niveau phonologique vs. réduction vocalique et consonantique dans les contextes [- accent]. En effet, la complexité syllabique et la réduction vocalique sont des facteurs qui contribuent à l'impression rythmique d'une langue.

Sur la base de cette compréhension du type rythmique, nous examinons la réduction vocalique et la chute de la voyelle finale (schwa) dans certains dialectes de la région de Naples et dans l'italien régional de Bari et de Naples. Les analyses ont été effectuées pour l' italien regional de Naples et de Bari sur le corpus AVIP (Archivio Varietà dell' Italiano Parlato, ftp://ftp.cirass.unina.it). Pour les dialectes de la région de Naples, les données analysées ont été enregistrées par l'auteur lors d'entretiens en tête à tête. Plus précisément, nous considérons les conséquences que la chute du schwa (qui conduit dans la plupart des cas à une syllabation inacceptable pour un locuteur italien) a sur la structure syllabique, en termes de localisation des ces langues sur le continuum 'isochronie syllabique – isochronie accentuelle'. Phonétiquement, le schwa est un son vocalique neutre, souvent considéré sans objectifs au niveau articulatoire (targetless, mais cf. Browman & Goldstein 1992). Les mesures spectro-acoustiques montrent que ses réalisations peuvent être caractérisée par un range variable par extension et forme spectrale, à partir de sons vocaliques multi-formantiques de durée considérable, jusqu'aux sons vocaliques résiduels qui ne rejoignent pas la cible (target), d'une simple pulsation glottale et énergie de résonance minimale (cf. Barry & Russo, 2002). Les structures syllabiques émergeant de la chute de la voyelle target (schwa final) montrent une deviation de ce qui est considéré acceptable en italien standard et étendent le range des types syllabiques "légitimes" en direction d'une classification rythmique typique des langues isoaccentuelles.

Barry, W.J. & Russo, M. (2002). Gradi di elisione dello schwa finale nelle varietà di Ischia e Pozzuoli (NA): un'analisi strumentale e implicazioni fonologiche. In: Regnicoli, A. (ed.). *XII^e Giornate di Studio del Gruppo di Fonetica Sperimentale*, Macerata 13-15 dicembre 2001. Il Calamo, Roma.

Browman, C.P. & Goldstein, L. (1992). Targetless schwa: an articulatory analysis. In: Docherty G. J. & Ladd, D. R. (eds): *Papers in Laboratory Phonology II. Gesture, Segment, Prosody*. Cambridge University Press, Cambridge, pp. 26-56.

Dauer, R.M. (1983). Stress-timing and syllable-timing reanalyzed. *Journal of Phonetics* 11, pp. 51-62.

E. Farnetani & S. Kori (1984). Effects of syllable and word structure on segmental durations in spoken Italian, *Quaderni del Centro di Studio per le Ricerche di Fonetica* 3, Padova, Progetto, pp. 143-187.

Roach, P. (1982). On the distinction between 'stress-timed' and 'syllable-timed' languages. In: Crystal, D. (ed.) *Linguistic controversies*. Arnold, London, pp. 73-79.

# IS PHONETICS A LAUGHING MATTER?

Jürgen Trouvain

Institute of Phonetics, Saarland University, Saarbrücken and

Phonetik-Büro Trouvain, Saarbrücken

trouvain@phonetik-buero.de

Laughter is a research subject for many disciplines including primatology, ethology, inter-cultural studies, medicine, neurobiology, conversational analysis, developmental psychology, and particularly emotion research. Can phonetics contribute to the research of this universal human pecularity of every-day communication?

In phonetic research, paralinguistic information in the acoustic communication channel is often identified with tone of voice, simultaneous with the verbal content of utterances, and only to a lesser extent with non-verbal vocalisations. Although laughter is one of the most important universal non-verbal vocalisations there are only a few studies performed by phoneticians on laughter. Since researchers from many disciplines have worked in the field, the terminology is heterogeneous and sometimes confusing. The missing common inventory for describing phonetic events makes data comparisons difficult and prevents the development of theories about the acoustic as well as the physiological and perceptual nature of laughter.

Examining the acoustic phonetic structure of a "typical" laugh reveals an alternating voiced-unvoiced pattern resembling a CV syllable structure (cf. the orthographic "haha" and "ahah") typically with an aspiration phase as consonant and a following vowel of varying quality. Moreover, the possible role of the inhalation phase with reverse phonation is usually ignored, although it is often strongly marked by a distinct increase in F0 and intensity.

Depending on the study, the vocalic segment is named "pulse", "note", "call", "event", "burst" but also contrary to phonetic terminology "syllable" or even "plosive". The consonantal segment is often seen as "interval" or "pause" rather than an aspiratory phase. The joint CV unit is referred to as "cycle", "syllable" or "call", enlarging the pool of terms. The inhalations sometimes found during laughter are seen as separators of different "bouts", and it often remains unclear whether a single laugh is exactly one bout or can consist of two or more bouts.

In addition to "typical" laughter, speech-synchronous forms such as speech-laughs and smiling while speaking can be distinguished from genuine laughter as well as faked, spoken/sung, and pathological forms and laughter due to tickling. Among the spontaneous and emotional forms of laughter, different types have been identified according to voicing and place of friction (oral or nasal). Not all studies agree on their findings regarding parameters related to F0, vowel quality, intensity, and duration. One reason for this can be found in the lack of common definitions of units and types of laughter.

The talk addresses the mentioned methodological issues by inspecting data containing laughter. This evokes a further methodological problem: How can we elicit spontaneous laughter? The inspection of speech data recorded for other purposes reveals that laughter is much more complex than expected from the classes and descriptions proposed so far.

Bachorowski, J.-A., Smoski, M.J. & Owren, M.J. (2001). The acoustic features of human laughter. *Journal of the Acoustical Society of America* 111, pp. 1582-1597.

Bickley, C. & Hunnicutt, S. (1992). Acoustic analysis of laughter. Proc. *Intern. Confer. of Spoken Language Processing ICSLP 92 Banff* (2), pp. 927-930.

Nwokah, E.E., Hsu, H.-C., Davies, P. & Fogel, A. (1999): The integration of laughter and speech in vocal communication: a dynamic systems perspective. *Journal of Speech, Language & Hearing Research* 42, pp. 880-894.

Ruch, W. & Ekman, P. (2001): The expressive pattern of laughter. In: Kaszniak, W.A. (ed) *Emotion, qualia, and consciousness.* Word Scientific Publisher, Tokyo, pp. 426-443.

Tartter, V.C. (1980): Happy talk: Perceptual and acoustic effects of smiling on speech. *Perception & Psychophysics* 27 (1), pp. 24-27.

Trouvain, J. (2001): Phonetic aspects of "speech-laughs". Proc. *Conference on Orality & Gestuality* (Orage 2001) Aix-en-Provence, pp. 634-639.

Trouvain, J. (2003). Segmenting phonetic units in laughter. To appear in Proc. *15th Int. Congress of Phonetic Sciences*.