Connected speech processes as multitier/multiarticulator prosodic modulations

Bernd Pompino-Marschall

ZAS – Centre for General Linguistics, Typology and Universals
Jägerstr. 10/11, D-10117 Berlin, Germany
bobby@zas.gwz-berlin.de

ABSTRACT

A model is proposed that interprets a variety of connected speech processes as resulting from prosodic modulations at different tiers of functional speech motor control along the hypo-hyper dimension [10]. The general background of the model is given by the trichotomy of A-, B- and C-prosodic phenomena [15] that together constitute the acoustic makeup of any speech utterance (with regard to their respective time domains at the uttarance/phrase level, the syllabic level and the segmental level).

1. Introduction

There are articulatory reorganizations in spontaneous speech that are not readily accounted for by assuming assimilatory processes alone. The German phrase

mit bunten Papierschlangen (‘with coloured paper-streamers’)

would be pronounced canonically – already partly reduced – as

[mɪtʰ ˈbʊntn̩ pʰaˈpʰiːˌɐˌʃlaŋn̩]

but in colloquial style, the following pronunciation is fairly regular [5]:

[mɪb̥ ˈbʊmpm̩ pʰaˈpʰiːˌɐˌʃlaŋː]

That assimilation is not enough to explain this reduced form can be seen if one compares this phrase with the following one:

mit runden Papierknödeln (‘with round paper-dumplings’);

here, no reorganization parallel to the above example is possible but only

[mɾʔ ˈʁʊn̩ˌm̩pʰaˈpʰiːˌɐˌknøːdl̩].

In this last example another quite regular process of spontaneous speech is seen, i.e. the use of glottal stop or glottalization instead of a coronal stop consonant (cf. [6, 7, 8]).

Also, the vocoid segments of the socalled ‘filled pauses’ – normally represented by [ə] – in spontaneous speech seem to differ in their spectral composition from the vowel nuclei of reduced syllables [14].

To cope with these different connected speech processes a quite general model of prosodic modulation is outlined in the rest of the paper.

2. Connected speech processes as prosodies

For our following discussion we will start from Tillmann's [15] characterization of the prosodic structure of speech utterances.

2.1. The homogenously prosodic nature of speech utterances

Perceptually, modulations of acoustic-auditory parameters as e.g. pitch, loudness, timbre etc. fall into three distinct categories depending on the frequency of modulation [15]: As long as we can follow the prosodic modulation as a variation of a distinct quality (e.g. the slow pitch variation of intonation) we may speak of A-prosodic phenomena. Changing to accellerated modulation, perception shifts from tracing a distinct quality to the impression of rhythm (B-prosody; e.g. the intensity modulation connected with syllable structure). At even higher modulatory frequencies secondary auditory qualities are the result (C-prosody). These perceptual qualities are normally described at the segmental level (e.g. trills, formant transitions 'coding' consonantal place of articulation).

In principle, speech utterances are all characterized by simultaneous prosodic modulations of the three types described above.

2.2. Prosodic modulations of complex speech signalparameters

The trichotomy of prosodic modulations sketched above not only holds for simple acoustic-auditory parameters but also for more complex auditory qualities resulting from different settings at different levels of speech motor control as e.g. 'voice quality' or 'tone of voice' [9]. We therefore propose a multitier/multiarticulator modelling of prosodic modulations in the broad sense as defined above.

One of the most general high-level prosodic tier may be characterized by the continuous modulation of the general adjustment of the vocal apparatus to speech (e.g. adducted/tense vocal chords, raised velum, mobile tongue body and blade) or to purely vegetative functioning (e.g. open/slack glottis, lowered velum, inactive tongue). This modulation of the most general setting of the vocal apparatus at the A-prosodic level seems to be responsible for the universally found phenomena of final lengthening (interpreted here as articulatory relaxation), F_0-declination as well as the aberrant acoustic structure of hesitation particles (being more neutral then reduced vowels, possibly nasalized). At the B-prosodic level this same modulation of 'articulatory tonicity' – triggered by metrical structure – may also be responsible for reductions in unstressed syllables and function words.

An independent second high-level prosodic modulation may manifest itself in the control of global speech rate (syllable rate as independent from intrinsically controlled articulatory speed). Quite a variety of connected speech processes seems to be a consequence of the interaction of both proposed prosodies: articulatory relaxation would also affect the strength of

interarticulatory timing (cf. e.g. [3]) and together with the enhancement of global tempo may result in changed timing relations between the gestures at different articulator-defined tiers (e.g. velar, laryngeal, labial, mandibular, coronal, dorsal) that show their intrinsically specified time constants. The wellknown gestural overlap phenomena – resulting in segmental 'quasi assimilations' and 'elisions' - as well as the instability in timing of e.g. laryngeal reflexes of a glottal stop before word-initial vowels or of internasal plosives at the segmental level [7, 8] and of syllabicity [12] could be interpreted as an interaction of the proposed multitier/multiarticulator prosodies.

3. Perspectives

3.1. What is needed?

To be able to develop a more formal model of prosodic control in connected speech – in the defined broad sense – it seems necessary to develop standards for multitiered annotation along the lines of e.g. [1, 2, 4, 13].

Although a very narrow transcription of spontaneous speech can show all these fine variations in pronunciation it misses to show the prosodic base of these processes.

3.2. An example

As a tentative example a short passage of spontaneous German (western mid-bavarian dialect[1]) together with annotations is shown in figure 1.[2]

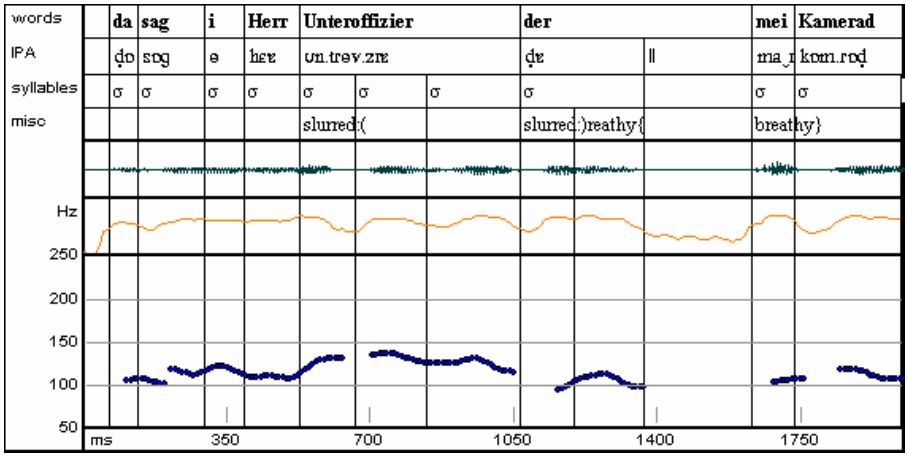

Figure 1: Orthographic, phonetic, syllabic and prosodic transcription of a short sample utterance (cont. next page)

[1] Speaker gg of [11] vol. 2, pp. 485ff.

[2] Generated with the help of the PitchWorks software of Scicon R&D (http://www.sciconrd.com).

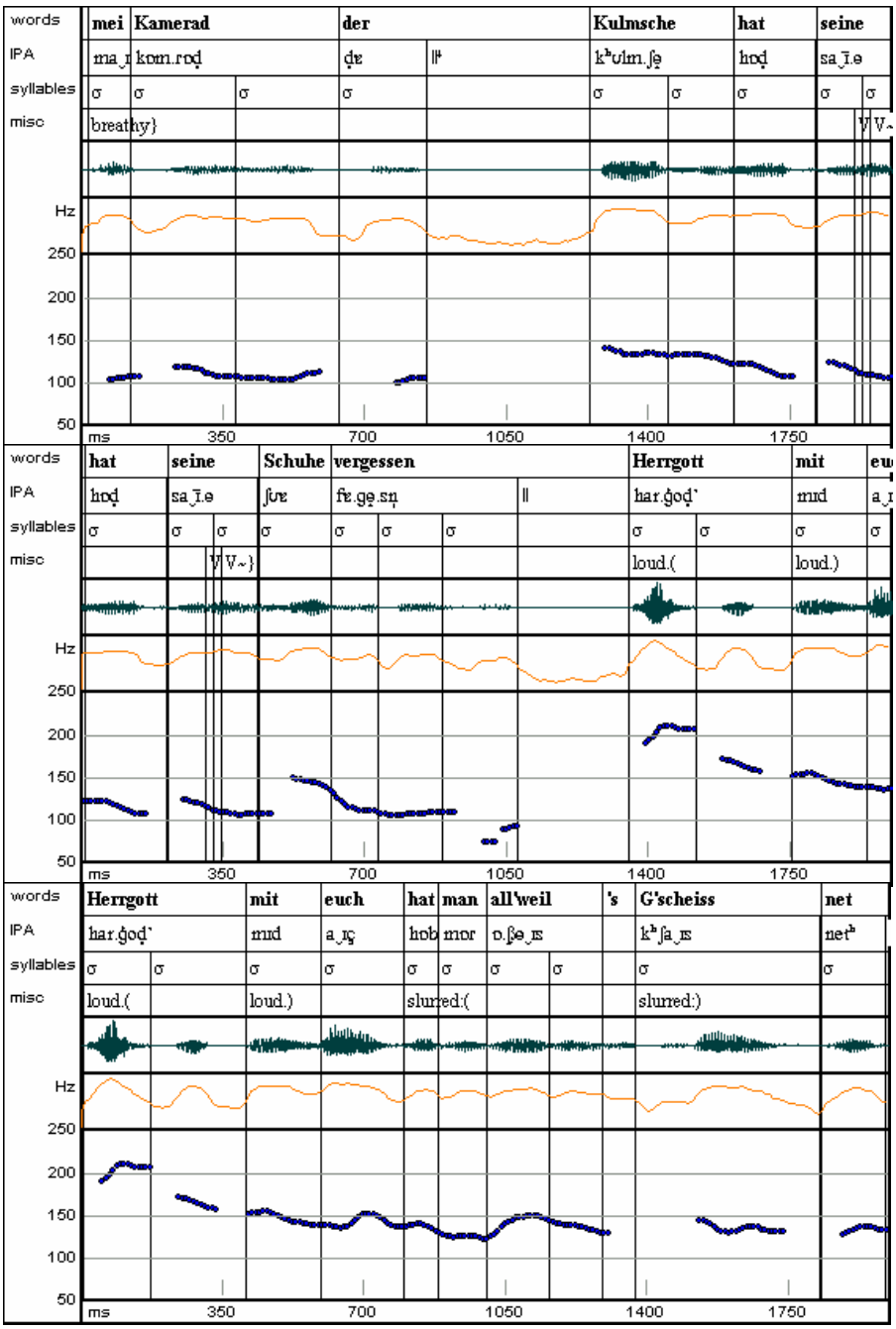

Figure 1 (continued)

(The translated utterance would be something like: “And so I said: “Sergant, my comrade, the one from Kulmbach, has forgotten his shoes, Sir!” “Damned, with you guys one is always in trouble!””)

ACKNOWLEDGEMENT

Work supported by German Research Council (DFG) Grant GWZ 4/5-1, P1.3 to the ZAS.

REFERENCES

[1] Ball, M. J.; J. Esling and C. Dickson (1995): The VoQS system for the transcription of voice quality. *Journal of the International Phonetic Association 25*, 71-80.

[2] Beckman, M. E. and G. A. Elam (1997): *Guidelines for ToBI labelling* (Version 3, March 1997). Ohio: Ohio State University Research Foundation.

[3] Browman, C. P. and L. M. Goldstein (1990), Tiers in articulatory phonology, with some implications for casual speech. In. J. Kinston and M. Beckman (eds.), *Papers in Laboratory Phonology I: Between the Grammar and the Physics of Speech*. Cambridge: Cambridge University Press, 341-376.

[4] Crystal, D. and R. Quirk (1964): *Systems of Prosodic and Paralinguistic Features in English*. The Hague: Mouton.

[5] Kohler, K. J. (1990): Segmental reduction in connected speech in German: phonological facts and phonetic explanations. In: W. J. Hardcastle and A. Marchal (eds.): *Speech Production and Speech Modelling*. Dordrecht: Kluver, 69-92.

[6] Kohler, K. J. (1994), Glottal stops and glottalization in German. *Phonetica 51*, 38-51.

[7] Kohler, K. J. (1999), Articulatory prosodies in German reduced speech. *Proceedings 14th International Congress of Phonetic Sciences, San Francisco*, 89-92.

[8] Kohler, K. J. (2000), Linguistic and paralinguistic functions of non-modal voice in connected speech. In: *Proceedings of the 5th Seminar on Speech Production: Models and Data,* Kloster Seeon, Bavaria, May 1st to 4th, 121-124.

[9] Laver, J. (1980), *The Phonetic Description of Voice Quality*. Cambridge: Cambridge University Press.

[10] Lindblom, B. (1990), Explaining phonetic variation: a sketch of the H&H theory. In: Hardcastle, W. J. Hardcastle and A. Marchal (eds.), *Speech Production and Speech Modelling*. Dordrecht: Kluver: 403-239.

[11] Pfeffer, J. A. and W. F. W. Lohnes (eds.) (1984), *Grunddeutsch: Texte zur gesprochenen deutschen Gegenwartssprache*. Tübingen: Niemeyer. 3 vols. [= Phonai 28-30].

[12] Pompino-Marschall, B. and P. Janker (1999), Production and perception of syllabic [n] in German. *Proceedings 14th International Congress of Phonetic Sciences, San Francisco*, 1079-1082.

[13] Roach, P., R. Stibbard, J. Osborne, S. Arnfield and J. Setter (1998), Transcription of prosodic and paralinguistic features of emotional speech. *Journal of the International Phonetic Association 28*, 83-94.

[14] Simpson, A., An acoustic analysis of hesitation particles in German. http://www.ipds.uni-kiel.de/forschung/hesitation.de.html.

[15] Tillmann, Hans G. (1980), *Phonetik. Lautsprachliche Zeichen, Sprachsignale und lautsprachlicher Kommunikationsprozeß*. Stuttgart: Klett-Cotta.