

LINGUISTIC AND PARALINGUISTIC FUNCTIONS OF NON-MODAL VOICE IN CONNECTED SPEECH

Klaus J. Kohler

Institute of Phonetics and Digital Speech Processing (IPDS), University of Kiel, Germany

Email: kk@ipds.uni-kiel.de

ABSTRACT

Non-modal voice (glottalization, breathiness, breathy voice) is examined with respect to four speech functions: vowel-related (word boundary), plosive-related, syllable-related (stød) and paralinguistic, utterance-related (phrase finality and truncation). The data are taken from 5 languages: German, English, Italian, French, Danish. They are analysed as phonatory prosodies in phrase-level phonetics and explained with reference to general principles of speech production.

1 INTRODUCTION

1.1 Definitions and linguistic functions of non-modal voice

In this paper, *non-modal voice* comprises various glottalization phenomena

- low frequency irregular glottal pulsing (variable in frequency, amplitude and waveform), in alternation with glottal stop, = *glottalization*
- *breathiness*
- *breathy voice*.

These glottalization phenomena fulfil a number of different linguistic functions in the languages of the world:

(a) *Vowel-related glottalization phenomena* signal the boundaries of words or morphemes beginning with vowels, typically in German, but also in other languages, e.g. English or French. The occurrence and phonetic manifestation of this function is also controlled by prosodic features, such as sentence accentuation, resulting in specific glottalization patterns for different languages (see Rodgers (1999) for German, and Dilley and Shattuck-Hufnagel (1995) for American English). The variation between glottal stop and any other glottalization phenomenon observed for this boundary signalling is associated with prosodic factors, e.g., degrees of sentence stress (cf. the link with ‘accent d’insistance’ in French), and with the degree of cohesion between successive lexical items.

(b) *Plosive-related glottalization phenomena* occur as reinforcement of plosives by a glottal stop or as replacement of plosives ranging from glottal stop to any other glottalization phenomenon, e.g. in German (Kohler, 1995, 1996a,b) or in English (Grice and Barry, 1991; Higginbottom, 1964; Pierrehumbert and Frisch, 1994; Roach, 1979).

(c) *Syllable-related glottalization phenomena* occur as characteristics of particular syllable types along a scale of phonatory weakening from glottal stop to glottalization, e.g. in the Danish stød (Fischer-Jørgensen, 1988, 1989). The “tense vs. lax larynx syndrome” set up for south-east Asian languages by Matisoff (1973) also relates to whole syllables, combining glottalization phenomena with prosodic and segmental features in phonological opposition.

(d) The *paralinguistic function of glottalization phenomena* manifests itself at the utterance level in two ways:

- as *laryngealization* in prosodic phrase-final *relaxation* of phonation, where the vocal folds prepare for abduction, and glottalization alternates with breathiness/breathy voice, not a glottal stop;

- as *truncation glottalization* in prosodic phrase-medial *tensing* of phonation at utterance breaks, where the vocal folds are adducted and glottalization alternates with a glottal stop (Nakatani and Hirschberg, 1994; Local and Kelly, 1986).

The question is whether these glottalization phenomena, which are controlled independently for these four linguistic functions but may nevertheless occur in the same utterances, are separate strands in speech production as sequential or superimposed events and whether they can then also be distinguished in speech perception and thereby serve communicative goals. In the German sentence “wir **könnten** ihn fragen” (“*we could ask him*”) there can be clearly distinct sequential vocal tract resonances for plosive and vowel-related glottalizations, contrasting with vowel-related glottalization alone in “wir können ihn fragen” (“*we can ask him*”). Similarly, the same two manifestations of glottalization can occur as temporally separated events in “achtzehnten **elften**” (“*18th November*”), where the glottalization for [ntn] is shifted left into the preceding vowel, and where the glottalization for the vowel [ɛ] occurs well inside it, thus ensuring a clear separation between them. Finally, in phrase-terminal “achtzehnten **Oktober**” (“*18th October*”), the sequencing of plosive and vowel-related glottalization as well as of phrase-final laryngealization can all be separable events; the former, although contiguous in time, are also distinguished spectrally and are auditorily attributable to two different functions. The above, as well as other examples discussed in this paper can be inspected as graphic signal and as sound output at the following URL:

www.ipds.uni-kiel.de/examples.html.

1.2 Databases for investigating glottalization phenomena

In the discussion that follows reference is made to databases and data descriptions of German, English, Italian, French and Danish.

For German, two large phonetically labelled acoustic corpora of read and spontaneous speech, compiled at IPDS Kiel, with 31,382 and 37,777 running regular lexical words (*The Kiel Corpus of Read /Spontaneous Speech*: IPDS, 1994, 1995, 1996, 1997a; Kohler, Pätzold, Simpson, 1995) were analysed within the *xassp* framework (IPDS, 1997b; Kohler, 1999b; Rodgers, 1999).

For American and British English, data descriptions by Dilley and Shattuck-Hufnagel (1995), Grice and Barry (1991), Higginbottom (1964), Pierrehumbert and Frisch (1994), and Roach (1979) were consulted.

For Italian, I looked at some of the spontaneous dialogues recorded in the MAPTASK scenario at Pisa and Napoli within the AVIP project (Archivio Varietà Italiano Parlato). The analysis was carried out in the *xassp* framework at Kiel. I am very grateful to Pier Marco Bertinetto and Federico Albano Leoni for making the recordings available to me.

For French, I looked at some spontaneous speech data, recorded by Angélique Amelot and Patricia Basset for their theses at the Phonetics Department of Paris III and kindly made available to me. The analysis was again carried out in the *xassp* framework at Kiel.

As regards these two Romance languages, the observations are only exemplary; they need subsequent systematic analyses, comparable to the ones done for German, on much broader databases with statistical evaluation.

The remarks on Danish are based on the very comprehensive description of the Danish stød by Eli Fischer-Jørgensen (1988, 1989).

2 GLOTTALIZATION DATA ACROSS LANGUAGES

2.1 German

2.1.1 Vowel-related glottalization. The marking of word and stem-initial vowel onset by a glottal stop is a time-honoured creed of German phonetics and phonology. It is only latterly, through the Kiel Corpus analysis (Rodgers, 1999), that it has been adjusted to get in line with the empirical basis of connected speech. The glottal stop ('hard onset') alternates with all other types of glottalization phenomena. These various non-modal phonation onsets on the whole occur irrespective of a sentence accent on the vowel. But if the word is accented, a glottal stop is more likely than other glottalization phenomena. The glottal stop is thus clearly a feature of reinforcement.

However, when the cohesion between words increases the junction feature may be weakened further and disappear altogether. This happens in cliticization of unaccented function words (e.g. personal pronouns in postverbal position, especially after auxiliaries, "hat er" ("has he"), or "und" in compound numerals, or pronouns after prepositions, "mit/von ihm/ihr..." ("with/by him/her")). This reduction may be such that an amplitude and/or an F0 dip remain as traces of the non-modal phonation onset, but complete levelling with modal voice is possible as well.

The occurrence of a glottal stop is also favoured in phrase-initial position, as against the preference of other glottalization phenomena phrase-medially.

2.1.2 Plosive-related glottalization. The phenomenon was first noticed in the acoustic and auditory processing and in the labelling of speech files within the Phondat and later-on the Verbmobil project (Kohler, Pätzold, Simpson, 1997). Several studies of the phenomenon have since been presented on the basis of the connected speech data from these two projects (Kohler, 1994, 1995, 1996a,b; Kohler and Rehor, 1996). The conditions of occurrence and the manifestation of plosive-related glottalization phenomena in read and spontaneous speech have recently been statistically evaluated in comprehensive database searches by Kohler (1999b), and can be summarized as follows:

(1) glottal stop and low-frequency glottal vibration general conditions

- A simple glottal valve action is used to cut off the air stream for stop articulation, added to, or instead of, a more complex combination of supraglottal oral/velic closures.
- The stop is not released into a vowel but is, in most cases, followed by another complete or partial oral occlusion - nasal, plosive or lateral.
- Irregular glottal pulsing instead of a glottal stop reduces, rather than blocks, the air stream. The frequency of this vibration differs from that of the (quasi)periodic environment; it is higher for tensing, lower for relaxation, the latter being typical.

specific contexts

(a) 'sonorant - plosive - sonorant' (especially nasal)

- Glottalization occurs for fortis and lenis stops at all places of articulation, word-internally or across word boundaries.
- The word-internal sequence results from [ə]-elision before nasals/laterals of canonical forms.

- The oral closure in nasal - plosive and plosive - nasal sequences is adjusted to the plosive place of articulation throughout.
- This oral closure is accompanied by velic opening as the complete or partial interruption of the air stream is transferred to the glottal valve.
- The following examples of low-frequency glottal vibration illustrate the various conditions: "könnten" [kœnn̩], "Lampen" [lamm̩], "halten" [ha]n], "Stunden" [ʃtund̩], "sind noch" [zɪn̩ nɔx], instead of the more elaborated canonical pronunciations [kœntʰən], [lampʰən], [haltʰən], [ʃtundən], [zɪnt nɔx]. Glottal stops are equally possible in these contexts.
- As regards the *timing* of glottalization there are four possible temporal alignments with the sonorant, e.g. /n/ in "könnten": medial, [n̩], is most common in all contexts final, [n̩], is next frequent for lenis stops initial, [n̩], is next frequent for fortis stops complete, [nn̩].
Glottalization may also extend, or be shifted, into the preceding vowel.
- With respect to the *duration* of glottalized nasal stretches, it has also been found to be quite variable (Kohler, 1996a). On the other hand, when glottalized sonorants were compared with sonorants that had modal-voice throughout, e.g. [kœnn̩] vs. [kœnn̩], the former turned out to be consistently longer. So it may be assumed that, from the production point of view, a canonical plosive unit can be represented in the signal by a duration trace as well, at least in systematically elicited lab speech data.

(b) 'vowel - (fricative) fortis plosive - consonant' (esp. nasal)

- There is a higher probability of glottally reinforced plosives (with velic raising) in this context than in context (a).
- In plosive - nasal sequences the same place adjustment occurs as in (a).
- In plosive - nasal sequences velic opening may occur very early after the oral occlusion, accompanied by glottal stop or low-frequency vibration.
- Examples for this context are *zweiten* [tsvai(?)n], *Leipzig* [lai? ptsɪç], *hat nicht* [hə? niç].

(2) breathiness and breathy voice

- In the complete nasal context of (a), voiceless or breathy-voiced nasals instead of plosives are also possible, breathy-voiced especially for lenis.
- This must be due to glottal (interarytenoidal) opening, which preserves the plosive phonation features, again combined with velic lowering, as required for the environment.
- Thus the modal-voice context of the nasal is still interrupted by different types of phonation, reflecting more complex plosive articulations.

(3) modal voice with(out) F0/amplitude modulation

- For lenis in the nasal context of (a), a further progression towards modal voice may be found, e.g. a reduction to [nn̩] in "einverstanden", "Stunden". For fortis this is only possible in unstressed function words, e.g. "könnten", and elements of compounds, e.g. "-zehnten" in numerals.
- This process may be complete, or there may be a weak trace of the plosive in the form of a medial amplitude and/or F0 dip in the nasal stretch. So the speaker can still signal a break, albeit towards the low effort end of a reduction scale ranging from plosive to complete nasalization.

2.1.3 Laryngealization and truncation glottalization. In phrase-final low-falling F0 patterns, modal-voice phonation quite regularly

turns into laryngealization, either low-frequency irregular vocal fold vibration ('creak' or 'creaky voice') or breathy voice or breath, often in sequence from creak (via breathy voice) to breath. These patterns indicate that the glottis progressively opens for breathing to mark the transition from speech to vegetative pulmonary action (Rogers, 1999).

When speakers interrupt their speech output because they have noticed an error in their production and want to correct it by resuming their sound generation at some earlier stage, or when they are interrupted by their dialogue partners and break off, they produce glottalization phenomena ranging from glottal stop to tight creak. In this truncation glottalization there is no alternation or sequencing with breathy voice or breath. So here the speaker does not prepare for exhalation, but, on the contrary, intends to hold his breath.

2.2 English

2.2.1 Vowel-related glottalization. In American and in British English the default word-initial vowel onset is non-glottalized. There may nevertheless be weak separation markers, 'internal open junctures', differentiating between, e.g., "a name" and "an aim" (Lehiste, 1960). Apart from durational characteristics, their phonetic exponents are amplitude dips. However, if a word is put in focus by strong sentence accent for contrast or emphasis, glottalization or, in the strongest case, a glottal stop may be used as the word separation marker. Glottalization also occurs as a 'spill-over' from laryngealization and truncation glottalization in 2.2.3.

2.2.2 Plosive-related glottalization. Glottalization, in alternation with, or in addition to, a more forceful glottal stop occurs as reinforcement of non-initial voiceless plosives or as manifestation of word-internal intervocalic /t/ in British English. In American English, the phenomenon has been recorded for voiceless stops in final (non-prevocalic) position and before sonorants and glides (nasals, laterals, /r/, /w/), including between nasals, e.g. "night-rate", "Clinton".

For the last-mentioned case, the textbooks and descriptions of British varieties of English state that the plosive is orally released and aspirated so that the condition for glottalization does not apply. More extensive searches of more comprehensive databases will have to be undertaken to confirm, or possibly disprove, this dictum. One thing is quite clear already that glottal stop or glottalization in an entirely nasal environment is also possible in British English if there is a word boundary before the right-hand nasal, as in "banknote", which is covered by the non-initial condition. However, cases like "spent an entire weekend" (with /ntn/) or "we lump them all together" (with /mpm/) seem to be possible as well.

2.2.3 Laryngealization and truncation glottalization. The same conditions for, and manifestations of, laryngealization and truncation glottalization seem to apply to English as to German. As a switch from speech production to breathing or as a mechanism to stop speech production, the processes must be assumed to be language independent.

2.3 The Romance languages Italian and French

2.3.1 Vowel-related glottalization. In both languages the default word-initial vowel onset is non-glottalized. Under strong sentence accent, glottalization phenomena appear, and phrase-final laryngealization and truncation glottalization are carried over to following word-initial vowels (see 2.3.3). Glottalization phenomena around word-initial vowels can thus become a cue to phrasing.

2.3.2 Plosive-related glottalization. Italian lacks the phonotactic structures required for the occurrence of this phenomenon: syllable-final plosives (except for such isolated cases as "Etna"). In French,

word-final plosives occur with the possibility of reinforcement to ejectives in utterance-final position. The other plosive contexts for glottalization in the Germanic languages also occur through [ə]-elision, as in "maintenant", but no cases of glottalization have been found so far. In this particular word, the tendency is towards modal nasal voice [nn] or even [n], comparable to the process found in German "Stunden" [nn]. This area requires more detailed analysis on the basis of a much larger database.

2.3.3 Laryngealization and truncation glottalization. The same patterns apply to these two Romance languages as to the two Germanic ones.

2.4 Danish

Glottalization alternating with reinforced glottal stop is also a common feature of the Danish stød. It is a prosodic phenomenon of certain syllable types, requiring a sufficiently long stretch of voicing, either long vowels or short vowels + sonorants, bound to primary or secondary stress. The prosodic nature is strengthened by the relation of the stød to accent 1 words in Norwegian and Swedish. The phenomenon can be illustrated by "vennen" (definite form of "ven" ("friend")), which is [vennɛn], besides emphatic [venʔn]. It coincides phonetically with the very common realizations of German *wenden* "to turn", having a glottalized nasal or glottal stop inside the nasal. So from the phonatory point of view German has a stød. And American English has one as well, for example in the pronunciation of the name *Fenton*. The distributions of this phonetic feature and their phonological functions are, of course, completely different.

How the general processes of phrase-final laryngealization and truncation glottalization interfere with, or overlay, the manifestation of the stød is at present unknown. This is a further field of investigation.

3 EXPLAINING THE DATA WITH REFERENCE TO GENERAL PRINCIPLES OF SPEECH PRODUCTION

3.1 Physiological explanation of glottalization phenomena

Physiological data on the Danish stød (Fischer-Jørgensen, 1988, 1989) show different degrees of medial vocal fold compression, due to increased/decreased vocalis and lateral crico-arytenoid muscle activities for glottal stop as against glottalization. So the glottal stop can be associated with a feature of reinforcement, compared with one of weakening in other glottalization phenomena. Extrapolating from these data of one particular language to the comparable alternation of glottal stop and non-modal phonation types in vowel-onset and plosive production of other languages we may hypothesize that the same strength relationship obtains there as well.

Phrase-final laryngealization ('creak' or 'creaky voice') and its alternation with breathy voice or breathiness fits in with low-F₀ utterance-final relaxation in preparation of glottal opening for the non-speech function of vegetative breathing; this relaxation excludes the glottal stop.

Utterance-internal speech truncation before correction is most effectively achieved by cutting off the air stream at the glottal valve; thus tensing of the vocal folds for a glottal stop would be the most natural process.

3.2 Economy of effort in plosive production

In both contextual patterns for plosive production in German, (a) and (b) in 2.1.2 (1), and similar patterns in English (2.2.2), two steps can be postulated for the simplification of the articulatory program:

- (1) the elimination of a central oral opening-closing gesture, resulting in nasal or lateral plosion,

- (II) the elimination of velic raising during the oral occlusion and the transfer of air stream control to the glottis, resulting in glottalization phenomena.

In a bilateral nasal context of pattern (a), (II) eliminates the need for a synchronization of velic control with other vocal tract components; the velum can remain lowered in the entire sequence, but the listener still receives a signal break for a canonical plosive through a glottal stop, glottalization or some other change in phonation. So articulatory reorganization takes place, which fulfils the same communicative function. Pattern (a) provides the most natural environment for articulatory economy through reorganization of stop production: glottal control can take over the function of air stream interference from velic control, allowing all other supraglottal settings - oral occlusion, place of articulation and velic lowering - to stay the same as needed for the nasal environment. Moreover, reducing the extent of velic movement and giving greater flexibility to its synchronization with other articulatory gestures accords very well with a sluggish articulator.

The temporal indeterminacy of velic action is further heightened by a temporal flexibility of phonation changes, as found in the extremely variable position and duration of glottalization within a sonorant (see 2.1.2). Although the duration of glottalized as against modal-voice sonorants can be longer this increase also varies a great deal from speaker to speaker. So in view of this enormous temporal variability in connected, and especially in spontaneous speech, it is doubtful whether the longer duration of glottalized sonorants is a generally established production pattern, and even more doubtful whether it plays an important role in speech communication.

In the short closure of lenis stops, this flexible timing of the velum triggers an even greater reduction of its movement, resulting in complete assimilation to modal-voice nasality (see 2.1.2.(3)). In all the modifications of 'plosive + nasal' syllables, found in the database, the demands on articulator coordination are reduced; this helps to ease production whenever called for by context of situation. In particular, unscripted dialogue speech, as against a reading style, sets a frame for greater articulatory imprecision; so there are more articulatory reductions of the types described under this condition.

3.3 Separation and interaction of the various strands of non-modal phonation

The data in the languages investigated, and the very extensive German ones in particular, show that the various strands of non-modal phonation, relating to four different speech functions, are separable.

But there are also interactions, especially between the phonatory manifestations of the paralinguistic and linguistic functions. This aspect has to be studied in more detail.

4 CONCLUSION

The data discussed in this paper have to a large extent escaped notice for a long time because phoneticians and phonologists have been preoccupied with citation form pronunciations of lexical entries and with their segmental representations. But the increasing interest in connected, more specifically spontaneous speech databases has made it mandatory for researchers to enter into the phonetics and phonology above the word in real-life communication, and it is in this domain that glottalization phenomena abound. In dealing with such phrase-level features it also becomes necessary to develop a new paradigm for speech analysis that moves away from a strictly linear phonemic framework towards a componential analysis to accommodate the temporal indeterminacies of speech production and perception successfully (Kohler, 1999a). Thus the four strands of non-modal voice (1.1)

are best treated as phonatory word (a), syllable (b,c) and utterance (d) prosodies, respectively.

REFERENCES

- Dilley, L. C., and Shattuck-Hufnagel, S. (1995). Variability in glottalization of word onset vowels in American English. In *Proc. XIIIth ICPHS*, Volume 4, Stockholm, pp. 586-589.
- Fischer-Jørgensen, E. (1988). Phonetic analysis of the stød in Standard Danish. *Phonetica* 46, 1-59.
- Fischer-Jørgensen, E. (1989). *A phonetic study of the stød in Standard Danish*. University of Turku Phonetics.
- Grice, M., and Barry, W. (1991). Problems of transcription and labelling in the specification of segmental and prosodic structure. In *Proc. XIIIth ICPHS*, Volume 5, Aix-en-Provence, pp. 66-69.
- Higginbottom, E. (1964). Glottal reinforcement in English. *Trans. Phil. Soc.*, 129-142.
- IPDS (1994, 1995, 1996, 1997a). *The Kiel Corpus of Read/Spontaneous Speech*. Kiel: IPDS. = Kiel CD-ROMs #1, #2, #3, #4.
- IPDS (1997b). *xassp user's manual (Advanced Speech Signal Processor under the X Window System)*. AIPUK 32, 31-115.
- Kohler, K. J. (1994). Glottal stops and glottalization in German. Data and theory of connected speech processes. *Phonetica* 51, 38-51.
- Kohler, K. J. (1995). The realization of plosives in nasal/lateral environments in spontaneous speech in German. In *Proc. XIIIth ICPHS*, Volume 2, Stockholm, pp. 210-213.
- Kohler, K. J. (1996a). Phonetic realization of German /ə/-syllables. *AIPUK* 30, 159-194.
- Kohler, K. J. (1996b). Glottal stop and glottalization - A prosody in European languages. *AIPUK* 30, 207-216.
- Kohler, K. J. (1999a). Articulatory prosodies in German reduced speech. In *Proc. XIVth ICPHS*, Volume 1, San Francisco, pp. 89-92.
- Kohler, K. J. (1999b). Plosive-related glottalization phenomena in read and spontaneous speech. A stød in German? *AIPUK* 34, 285-325.
- Kohler, K. J., Pätzold, M., Simpson, A. (1995). *From scenario to segment - The controlled elicitation, transcription, segmentation and labelling of spontaneous speech*. *AIPUK* 29, 1-141.
- Kohler, K. J., and Rehor, C. (1996). Glottalization across word and syllable boundaries. *AIPUK* 30, 195-206.
- Lehiste, I. (1960). An acoustic-phonetic study of internal open juncture. *Phonetica* 5 (Supplement), 1-54.
- Local, J. and J. Kelly (1986). Projection and 'silences': notes on phonetic detail and conversational structure. *Human Studies* 9, 185-204.
- Matisoff, J. A. (1973). Tonogenesis in Southeast Asia. In: L.M. Hyman (Ed.), *Consonant Types & Tones. Southern California Occasional Papers in Linguistics*, Volume 1, pp. 71-95, Los Angeles: Linguistics Program University of Southern California.
- Nakatani, C. and J. Hirschberg (1994). A corpus-based study of repair cues in spontaneous speech. *JASA* 95, 1603-1616.
- Pierrehumbert, J. and Frisch, S. (1994). Source allophony and speech synthesis. In *Proc. ESCA/IEEE Workshop on Speech Synthesis*, pp. 1-4.
- Roach, P. (1979). Laryngeal-oral coarticulation in glottalized English plosives. *JIPA* 9(1), 2-6.
- Rodgers, J. (1999). Three influences on glottalization in read and spontaneous German speech. *AIPUK* 34, 178-284.