

## PHONETICS AND PHONOLOGY OF SPEAKING STYLES

W. J. Barry

*Institute of Phonetics, University of the Saarland, Germany*

### ABSTRACT

This discussion attempts to identify some questions which have been neglected in the recent revived interest in speaking styles, and to discuss the theoretical implications of incorporating them.

### THE SCOPE OF SPEAKING STYLES

#### Present approaches

In a recent paper on trends in speaking-style research, Maxine Eskanazi [1] selected three dimensions along which styles may be located.

- the degree of (presumably phonetic) intelligibility required
- familiarity between speaker and audience, and
- the social strata of those speaking.

She acknowledges the limiting nature of such a three-dimensional framework with a comment on the limits of the data available, and summarises a more comprehensive view of style with a complex of factors seen as operating mainly on the speaker:

"Style reflects the action of the environment upon the individual and the individual upon the environment. It is his perception of the various "status" levels of his listener and of the type of situation in which he finds himself. It is also the projection of himself, his background, and is a setting of the type and tone of conversation he wishes to have. All of this is a mixture of conscious and unconscious (...) effort on his part and is not always perceived in the same way as it is intended" ([1] p. 502)

The limits of the data available to re-define the framework, a scan of the literature suggests, are due to the scientific need for controls in data acquisition if differences in the results of analysis are to be attributed to the postulated differences in style. There is an understandably disproportionate emphasis on the difference between "read" or "prepared" speech and "spontaneous" speech. Another axis along which

analytic comparisons are frequently made is "clear" vs. "casual" speech. Both these axes dominate, I suggest, because they can be turned into relatively easy instructions to, or situational parameters for the speakers during the acquisition phase. The effect on the three-dimensional framework suggested by Eskanazi is practically to collapse it into a one-dimensional "formal - casual" axis.

As an aside, it might be added that there are non-scientific interest-driven reasons for the restriction of speech types (I purposely avoid the term speaking style) of which we are all aware. It is therefore all the more important to ask whether there are convincing scientific reasons for *not* expanding the scope of speaking-style research.

My first question is, therefore, whether there is not an undue reduction of the concept of speaking style under discussion, and whether there are not important basic aspects which are not covered by the suggested descriptive framework?

Within this symposium, an extension of scope attempted in this paper is not intended as a focus for the other papers, but as a backdrop against which to place them in the perspective of the overall goal of characterising speaking style.

#### Other Dimensions of Speaking Style

An important consideration is that, in contrast to the production-oriented dimensions for categorising speaking styles summarised above, the concept of speaking style is fundamentally listener-oriented. Whether we consider the expressions used to describe someone's style of speaking, or the schooling of speakers for a particular task (politicians, managers, salespeople, etc.), it is the effect on the listener which is primary. The character actor is the epitome of audience orientation.

Two assumptions are implicit in the term "speaking style", first that listeners have a personal neutral baseline against

which they judge individual utterances, second that speakers' styles deviate to a greater or lesser degree as a product of their states of mind and their communicative intentions from their individual production baselines. If speaking styles are seen as communicatively significant, a considerable degree of isomorphism must be assumed between the production and listener baselines and the dimensions of variation around them.

Defining speech-production tasks and locating the resultant speech within the descriptive framework of the kind suggested in [1] does not allow for anomalous speaker performance in terms of effects on listeners. In particular the use of production categories such as "reading" and "spontaneous" as a basis for speaking-style analysis is misleading. The two categories cut across the dimensions of the Eskanazi framework, and conflate a vast number of different production styles. Radically differing performance at reading aloud, and a wide variety of situational and task variants within the category "spontaneous speech" make quantitatively meaningful statements across studies impossible.

For Speech Technology purposes, a differentiation of read and non-read speech may appear desirable, but it is not only very restricting in the search for phonetic correlates of communicatively relevant style features, it may well lead to nothing for Speech Technology. In most individuals, and in most groups within a given task definition, the difference may well be statistically clear, but the differences are not likely to be generalisable.

#### Phonetic Analysis

The dimensions of phonetic analysis, on the other hand, are common across speakers and tasks. They need to cover the three basic aspects of speech production: a) laryngeal excitation characteristics and related aspects of voice quality, b) modifications to the segmental structure of words and word sequences, and c) rhythmic and tonal properties of utterances resulting from the rate of speech, the grouping of words, and from the accentual and intonational patterns used.

Of these three areas of analysis, the second and third have received almost

exclusive attention in recent work on speaking styles, as the other contributions in this symposium document. Voice quality has presumably been largely neglected for technical rather than theoretical reasons. Its contribution to the description of speakers and interaction in discourse are long established under the term paralinguistic [2], and would add a valuable dimension to speaking-style analyses (cf. Campbell, this symposium), though its role as a functional prosodic feature also needs formal delimitation (at least in southern British English voice-quality alone ([±breathy]) can turn a rising contradictory query into a disbelieving query in the "He's not(?)" response to: "The bishop's coming to tea today.")

#### Practical Applications

The three phonetic analysis dimensions mentioned above form the basis of speaker analysis in the present-day, phonetically oriented approach to forensic speaker identification, though the short-term situational and long-term speaker factors are conflated. As a first step, forensic phonetic work calls for the detailed auditory and instrumental analysis of any speech samples produced by the perpetrators, to give a profile of the person(s) being sought, before any suspects are asked to produce speech samples for comparison [3].

It is the listener-oriented side of speaking style that attracts the scrutiny of, and is modified by the image-building consultants engaged by politicians and other public figures. The consultants may not have a battery of phonetic facts available, but they are obviously able to identify, and communicate impressionistically, the parameters they think should be changed. Scientific phonetic goals should not be confused with PR (though commercially-oriented funding pressure has already taken its toll), but there is obviously an area of important practical application that mainstream phonetic research is neglecting at present.

Even in speech technology, the demands of which are to a large extent responsible for the restricted scope of work on speaking style in the past two decades, an open-minded consideration of the factors influencing speakers' production, once they depart from the

level of single-word utterances, is vital for open-ended progress. The implications for ASR are naturally greater than for speech synthesis because the recogniser in public service has to face the full range of voice-types and of personality-driven and situationally caused speaking styles. But in synthesis too, without quantitative knowledge of the phonetic parameters and the phonetic-phonological processes involved, progress towards task-differentiated synthesis will be blocked (see Campbell, this symposium).

The impact of these areas of professional activity on modern life is to my mind a sufficient reason for extending controlled study of speaking style beyond the formal-casual axis. This is by no means to say that the effects along that axis have been sufficiently explained, but it would, for example, be a disservice to engineers working in speech technology, to imply that the problems of spontaneous-speech will be solved by systematising that axis. To phoneticians, the vast complexity of speech production phenomena and their sensitivity to all aspects of the social, situational and psychological scenario goes (almost) without saying. But, as comments from non-linguists frequently document, to them speech is still the audible part of what can be written on a sheet of paper, and even for many engineers working in speech technology a lot of effort has to be invested to show that signal variability can be related to systematic dimensions of the communicative situation, not just to the phonological, morphological and syntactic structure.

#### Difficulties of Analysis

The arguments for extending the classification base of speaking-style studies may convince without changing the unfeasibility of actually doing it. Past phonetic studies of emotional and attitudinal aspects of speech, as judged by listeners, are indeed not easily translatable into general statements of phonetic and/or phonological structure [4-9]. They appear to be much less easily controlled than speaker and production variables.

This is no doubt to some extent a valid objection, because if listener-oriented

categories of speaking style are to be studied, control, in the form of listener-group judgements, has to follow the selection of a variety of spoken texts by each speaker to be analysed. As for the data collection itself, care has to be taken, a) to create a speech-production task and situation where the parameters under scrutiny are liable to occur, b) to obtain the validation by the speakers of the attitude, emotion, intention etc. underlying a particular utterance.

However, given these steps, control is in fact greater, due to formalised listener-group acceptance of the assumed category, than in solely production-oriented experiments. There, control is limited to instructions to the speakers (e.g. read, read/speak clearly, etc.) or definition of the interactional situation. The communicative effect of a speaker's behaviour in terms of a listener-group's descriptive categories is rarely formally checked.

#### PHONETICS AND PHONOLOGY

Another advantage of an extended scope of speaking style, though completely intrinsic to Phonetics and Phonology, would be a broader frame for the discussion of a topical theoretical issue, namely the question of phonetic vs. phonology within speaking-styles, with special reference to reduction phenomena.

To avoid misunderstanding, let me define my understanding of the part of phonology implicated in this discussion. I am considering the systematic basis of speech production at the level of motor programming, not merely a formal meta-system for the representation of parts of utterances (although that level of description is both necessary and important as an interface to morphology and syntax, as is a consideration of perception in phonology). The articulatorily based approach by Browman & Goldstein (B&G) [10, 11] exemplifies the aspect of phonology focussed on here. Their claims are currently under discussion and are eminently relevant to a discussion of speaking styles.

Briefly: B&G advocate a structure of gestures as phonological units which overlap and may be reduced, but which are not exchanged for other gestural units. Thus, all reduced forms have the

same phonological form, and all reductions (e.g. along the formal-casual continuum) must be regarded as phonetic processes. Others (e.g. Holst & Nolan [12], Kohler [13]) while not arguing against a wide range of phonetic forms which can be explained as stages along a continuum of articulatory reduction (which may or may not be represented as different sequences of phonetic segments), argue for an ultimate form in the range of observed variants which reflects a categorical shift in the underlying phonological structure (see also Kohler, this symposium).

As so often in academic issues of principle, neither standpoint is provable because the ultimate answer is inaccessible to observation. In this case it lies in the production plans of the speaker. Nolan's and Kohler's indirect evidence in the form of careful analysis of surface behaviour is convincing from their own standpoint (see Kohler's extreme case of anticipatory assimilation "mit bunten Papierschlängen", this symposium), but apparently fails to convince the opposing camp.

The crux seems to be the degree of overlap and the degree of reduction that B&G's model allows, ultimately, how abstract their initially articulatorily interpretable model has become. However, that is no indication of what "gestural phonology" natural speakers actually possess, if gestures are what the organisational units of the production plan happen to be

#### Phonological Switch in Speaking Styles

Although not intended, the connection between reduction processes and the formal-casual speaking style axis can imply the general lack of phonological change with style change; i.e. style is a surface phonetic phenomenon.

Consideration of a wider range of stylistic phenomena shows, however, that changes only interpretable as "phonological" switches as well as changes in the surface-phonetic form are commonplace. Since many of these switches can be observed in connection with the same sort of socio-communicative variables as are implicated in the three dimensions behind the formal-casual continuum, the suggestion is, that

phonological switches of the "gestural reorganisation" type are equally plausible.

#### Phonological Switches in Prosody

Before considering possible phonological switches involving segmental restructuring at the level of word sequences, we can illustrate that switches are commonplace at higher levels of phonological structure.

There are formalised speaking roles with recognized (and immediately recognizable) intonational patterns which differ radically from those used in any other form of speech. Examples of these, admittedly extreme in some cases, are certainly horse-race commentators, auctioneers, marketenders, sermons and community praying. The latter two cases have been described for English by Crystal [2].

Horse-race commentaries have not, to my knowledge, been described systematically, but the pattern is presumably familiar to many people; it seems to be similar across a number of languages: In British English there is a clear monotone-oriented rule, with definite, race-stage-oriented resets (to a high pitch) with tempo and volume increases from one series of "intonation units" to the next, and with a sudden *rallentando* and *decrescendo* combined with a short series of resets to a lower pitch and a final low falling contour from the moment the winning horse finishes.

The other examples can be similarly characterised with a combination of local prosodic rules and an "operation-bound" pattern of intonation-unit sequences. The switch into any of these intonational systems is of course strictly situation-bound.

Another example of intonational switch, of a more subtle kind, has been observed during intonation work in Saarbrücken with southeast Italian (Bari) speakers [14]. In order to elicit spontaneous speech, a map-task is given [15, 16] in which one person verbally guides another along a route. Certain differences exist between the two maps, leading to frequent requests for information, confirmation etc. The two speakers "play" the game twice, once as guide, once being guided. The speakers

are unaware of the purpose of the recordings being made; only after the event do they fill in a form asking them about their language (dialect) habits. In one case, one of the speakers saw the questionnaire between the two "games". The researchers monitoring the recordings were puzzled that the speaker, who had been asking questions with a typical Bari rise + fall intonation [14] during the first game, consistently used the standard Italian rising contour in the second. When questioned about it, the speaker was in fact unaware of the manner of the "style" change, but confirmed that he had realized about the interest in accent from the questionnaire. This had triggered the unconscious switch to the more socially acceptable form.

A similar switch was observed in Palermo Italian in an earlier study in which dialect forms of question intonation were first elicited in a "game" situation and then presented in written form for controlled reading. The speakers again switched to the standard Italian intonation, and were unable to read with the dialect intonation until the questions were massively contextualized with dialectal precursors ([16] p. 143).

### Segmental Switches

Both of the above examples are long-term switches; i.e. a phonological variant is selected which the speaker uses in a given situation. However, shorter-term switches are also common. It is extremely common for bilinguals to switch from one language to the other, and often back again, in the middle of a sentence. What is more, the switch is often unconscious, and the speaker can continue some time without being aware of having switched. This well-studied phenomena [18, 19] is always linked to a "trigger" word or event which is associated with the language that is switched into. This might be considered too far removed from speaking-style phenomena; after all, a sentence, or part of one, requires morphological and syntactic switches as well as phonological and phonetic ones. However, a minimal switch may extend for one word only, and the situational factors facilitating switches are likely to be located at the familiar end of the familiar - non-familiar dimension.

The examples so far indicate that, given the right situational factors, individuals generate utterances while selecting from different (parallel) sets of phonological rules.

Reducing the bilinguality to bi-dialectality, we get ever closer to normal speakers and their phonetic-phonological adjustment to situational factors. Presumably, most people have experienced dialect speakers' in a standard-language situation drifting into dialect as they discuss a point among themselves. Among highly educated people, who possibly spend more time speaking "standard" than dialect, it is a signal of solidarity and familiarity (even complicity, since it is often employed when negotiating a favour or service).

Important for the discussion of the "phonological switch" phenomenon is the considered use, above, of the term "drifting into dialect". My observation of Saarland dialect speakers, confirmed by phonetically aware dialect speakers (in fact the phenomenon appears to be general to bi-dialectal areas) is that the switch is seldom from standard German straight into full dialect. There is an intermediate form which is clearly close to the standard in lexical choice and even syntax but contains definite phonological and/or morpho-phonological modifications:

Standard	-->	Saarbrücken
/...st/	-->	/...ʃt/
(weiß du	-->	[vɛ:ʃtə])
(erst	-->	[ɛʃt])
/..V(:)Cən/	-->	/..V(:)Cə/ or /..V(:)C/
(willst du messen	-->	[vɪlʃtə mɛs(ə)])

The use of standard words with a modified phonology is again important evidence for stylistically triggered phonological change. These alternative forms obviously cannot be explained by "reduction" processes. The speakers are either "switching on" a phonological rule which carries out the change to the dialectal sound structure, or they have both standard and dialectal forms in their lexicon to access appropriately, depending on the situation.

Finally, we find that non-bilingual, non-bidialectal speakers produce forms

of the same words and phrases which are far apart on a putative continuum of change, possibly from near the maximum form to near the most reduced, but do not produce forms in between. For example:

"going to" (/ˈgəʊɪŋ tu:/)  
can easily be shown to reduce to

"gonna" (/ˈɡʌnə/)  
in a logical series of steps. However, the segmentally very reduced form is likely to appear at relatively slow speech rates, and even in accented position. This rules out the possibility of it emerging as a result of overlapping underlying gestures. The alternative, stylistically marked lexical entry may therefore be a more plausible explanation.

Similar examples are those given by Kohler (this symposium) as "stereotype formulae", which can also be interpreted as implying its own lexical (sub-) entry: 'kyou (/kju:/) for "thank you"

'nAbend (/ˈnɑ:bmt/) for "guten Abend"  
In both cases a lexically stressed syllable is missing compared to the standard expression, which cannot be attributed to any normal elision rule.

Both "thank you" and "guten Abend" are written as two words but are presumably stored as single entries because of their formulaic function (stroke patients, unable to articulate more than single monyllabic words, are known to produce formulaic expressions such as "thank you" spontaneously).

### Lexical Entry or Gestural Reducing?

Explaining the phonological switch only in terms of alternative lexical access rather than an on-line change of phonological structuring during the articulatory planning stage of production is not satisfactory. The parallel between these examples and the critical issue of gestural overlap or phonological recoding only applies when the reduction is within a single lexical item. Words like: *greatcoat*: /ɡreɪtkəʊt/ <--> /ɡreɪkkəʊt/ and *handbag*: /hændbæg/ <--> /hæmbbæg/ satisfy this, but the similarity between this intra-word alternation and the possibility of the same alternation occurring across word boundaries, e.g:

*He pulled his hand back*:  
/... hænd bæk/ <--> /... hæm: bæk/

makes an on-line phonological recoding mechanism rather than just style-based alternatives in the lexicon necessary, unless B&G can convince the other side of the plausibility of an absolute overlap + reduction principle.

### SUMMARY

In the first section of this contribution to the speaking style discussion, we offered arguments for a need to extend the scope in the study of speaking styles, and argued both that the listener-orientation is basic to the concept of speaking-style, and that it has been largely ignored in recent work.

It was also argued that scientific stringency need not suffer and that the extended scope would, in fact, provide a step towards the refinement of the framework which Eskanzi considers necessary ([1] p. 501).

In the second section, we discussed the phonetics-phonology issue of gestural overlap and reduction vs. phonological reorganisation of gestures within the suggested wider scope of speaking style work. The commonplace occurrence of style-linked phonological switches was illustrated at different levels.

In the case of segmental variants, the alternative lexicon entries were seen as a possible explanation for single-word cases. Across word boundaries, however, similar variants could not be explained in this way, and style-based phonological processes were implicated. It was seen, though, that parallels with the original gestural issue were limited and highlighted it rather than resolving it.

### REFERENCES

- [1] Eskanzi, M. (1993): Trends in speaking styles research. *Proceedings Eurospeech 93*, Berlin. 501-509
- [2] Crystal, D. (1975): *The English Tone of Voice. Essays in intonation, prosody and paralanguage*, London: Arnold
- [3] Künzel, H.J. (1987): *Sprechereerkennung. Grundzüge forensischer Sprachverarbeitung*. Heidelberg: Kriminalistik Verlag
- [4] Fónagy, I. (1964): L'information du style verbal. *Linguistics* 4, 19-47
- [5] Fónagy, I. (1969): Métaphore d'intonation et changement d'intonation.

*Bulletin Soc. Linguistique de Paris* 64, 22-42

[6] Fónagy, I. & Bérard, E. (1972): "Il est huit heures": Contribution à l'analyse sémantique de la vive voix. *Phonetica* 26, 157-192

[7] Uldall, E.T. (1964): Attitudinal meaning conveyed by intonation contours. *Language & Speech* 3, 223-234

[8] Uldall, E. (1964): Dimensions of meaning in intonation. In: D. Abercrombie, D.B. Fry, P.A.D. MacCarthy, N.C. Scott & J.L.M. Trim (eds.), *In Honour of Daniel Jones*, London: Longmans, 271-279

[9] Hadding-Koch, K. & Studdert-Kennedy, M. (1974): Are you asking me, telling me, or talking to yourself? *J. Phonetics* 2, 7-14

[10] Browman, C.P. & Goldstein, L. (1992a): "Targetless" schwa: an articulatory analysis. In: G. J. Docherty and D. R. Ladd (Eds.), *Gesture, Segment, Prosody. Papers in Laboratory Phonology II*. Chapter 2, pp. 26-55. Cambridge: University Press.

[11] Browman, C.P. & Goldstein, L. (1992b): Articulatory Phonology: An Overview. *Phonetica* 49, 155-180.

[12] Holst, T. & Nolan, F. (in press): The influence of syntactic structures on [s] to [ʃ] assimilation. In: B. Connell and A. Arvaniti (eds.), *Phonology and Phonetic Evidence. Papers in Laboratory Phonology IV*. Cambridge: CUP.

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

[14] Grice, M., Savino, M., (1995): Low tone target versus "sag" in Bari Italian intonation; a perceptual experiment. *Proc. XIII Int. Cong. Phonetic Sciences*

[15] Grice, M., Benzmueller, R., Savino, M., Andreeva, B. (1995): The intonation of queries and checks across languages: Data from Map Task dialogues. *Proc. XIII Int. Cong. Phonetic Sciences*

[16] Anderson et al. (1991): The HCRC Map Task Corpus, *Language and Speech* 34, 351-366

[17] Grice, M. (1995): *The intonation of interrogation in Palermo Italian. Implications for intonation theory*. Niemeyer (=Linguistische Arbeiten 334)

[18] Clyne, M.G. (1967): *Transference and Triggering*. The Hague

[19] Clyne, M.G. (1969): Switching between language systems. *Actes du Xe Congrès International des Linguistes*. Bucarest, 343-349