

Trends in Speaking Styles Research

Maxine Eskénazi

LIMSI-CNRS, BP 133, 91403-ORSAY CEDEX, FRANCE

INTRODUCTION

One of the principal interests of researchers in automatic speech processing at present is the adequate representation of the great variability in the speech signal. Some search for common behaviour among many speakers, "invariants" in the signal, yet others search for a chart of all that varies in order to better understand the limits of variability. In a first approximation, the study of speaking styles should be of little interest to the first group. It would seem to have little effect on the nature of the linguistic message. Yet, when applying speech recognition systems to increasingly natural speech and realistic applications, it has been found that a change in speaking style affects recognition results. Elsewhere, in working with the hard-of-hearing, it was found that processors on hearing aids could be greatly improved if the information used when one speaks clearly is modelled. And yet elsewhere, developers of text-to-speech synthesis systems recognise that their systems could benefit in intelligibility and naturalness from models of clear speech. The characterisation and subsequent modelling of this style phenomenon therefore have become of great interest. As for the second group, taking a view that we must understand all sources of variability in order to correctly process speech [LR84] [LR90], style is one of the many dimensions to be explored to the limits of speaker space.

The term "speaking styles" has been used and defined in many ways [L192], almost as numerous as the authors who have dealt with the subject. It can be said that many suprasegmental changes that could not be attributed to a more well-defined area, such as emotion [MU93] have been thrown on the heap which has until present been called style, by the speech community [EM92]. We shall return to the definition of speaking style below. First it is necessary to see what has been accomplished in communities which have a longer experience in studying style, sociolinguistics and psycholinguistics.

PIONEERING STUDIES IN SOCIOLINGUISTICS AND PSYCHOLINGUISTICS

Based on large quantities of data in some cases, on personal experience in others, researchers in psycholinguistics have put down several building blocks which should serve as a point of departure for the speech community in its nascent undertakings.

The work of Labov [LA64] [LA72] put forward the idea that style changes as a function of the attention that the speaker

pays to his discourse (more attention furnishing a better-articulated, more formal style). He also postulated that the cause of this change was the social level at which the discussion was taking place. In one given conversation, however, he found that there are constant variations of style. For Joos [JO68] and Zwicky [ZW72], styles may be defined along an axis of casualness. For Zwicky, four degrees of casualness can be defined whereas Joos finds 5 (intimate, casual, consultative, formal, frozen). The one-dimensional aspect and fixed levels of these two authors were a first effort at organising various speaking styles, although they may seem too rigid by today's standards.

The above authors also listed the elements in speech which they observed which contributed to an impression of style change. These elements for Labov were of suprasegmental, quantifiable nature: change in tempo, in pitch, in intensity, in breathing, or presence of laughter. For Joos, they were of a syntactic nature: ellipsis, slang, absence of an article or a subject at the beginning of a sentence. Zwicky saw cues as contributing to either ease or brevity of articulation, since speed and the resulting articulation were the only elements of style change from his point of view. For ease, he listed assimilation, neutralisation, insertion, and transitional sounds. For brevity, he found geminate simplification, vowel contraction, deletion of "weakly articulated" segments, and monophthongisation. All of these authors admitted that style changed continually during a given discourse, theme to later be taken up by the work of Lindblom [LB90], [LB91], for example, and others. All of their views also take at least parts of the situation of the discourse (speaker-context-listener) into account.

DEFINING SPEAKING STYLES

In order to explore and relate the work on speaking styles in the literature, we shall now attempt to give the term, "speaking style" a data-driven definition. In doing so we are taking into account theoretical and quantitative findings in the literature. This definition also must take into account the most widely-used concepts of human communication, that is the message is sent from a speaker to a listener through a medium (channel) [JK76]. The definition below remains limited, due to the limits of the data on which it is based. The dimensions overlap, and account for many different factors of varying levels of complexity. Further work will certainly refine the axes and take others, which have not yet been explored, into account.

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 a 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 (voluntary/involuntary) effort on his part and is not always perceived in the same way as it was intended.

Changes in speaking styles take place often within a same conversation, the degree of attention to the clarity of discourse changing constantly [LA72], [[LB90]. A change in speaking may, according to our definition, be caused by a change in the self-image that the speaker wishes to project, in the type of information to be communicated, in the situation in which the interlocutors find themselves (including background noise, arrival or departure of other persons, the dialogue context), and in the image that the speaker has of the listener (can't hear well, personal background, etc.).

Different speakers, as we will see below, may use different ways to express a speaking style which is perceived to be the same; and different listeners may vary in the degree to which they perceive a speaking style. Some styles may be easier than others for some people to express, and for some to perceive.

This definition narrows the domain of style. Such variations as soft/loud, and probably fast and slow, for example, become subservient elements, used along with others, to create the impression of a given speaking style. The carriers of style are of segmental (stop articulation, or devoicing, for example) and suprasegmental (intensity, or F0 range, for example) nature.

In order to better capture the nature of speaking styles, it is useful to define the dimensions along which styles may be located. We have chosen to define three axes, compatible with the above definition.

Intelligibility-oriented

The degree of clarity that the speaker intends his message to have is the first dimension. It goes from minimum effort to be clear, when the message can be heard and understood well, to much effort when the channel is in some way noisy, or the listener has a problem understanding. This takes into account characteristics of the environment as well as the

listener, and the judgement here concerns elements of a more physical nature.

Familiarity

The expression of style may change greatly according to the speaker's familiarity with the listener. Extremes may go from identical twins to talking to someone from another culture who has little knowledge of the speaker's language and culture. The judgement here is based on more abstract elements, and takes the image of the listener in the dialogue context into account.

Social strata

The degree of cultivation that the speaker and listener wish to accord their conversation (see Labov) is the third axis. This goes from a totally colloquial and/or "lower class" tone to a "highly cultivated" and/or "upper class" tone. It takes into account the context in which the conversation is taking place as well as the backgrounds of the two interlocutors at an abstract level. It should be noted here, that in a less limited style space, the interaction of the speaker and his environment must be taken into account, either on this third axis, or on another. Since the literature has not, to our knowledge, yet dealt with this aspect, it is not included here.

Figure 1 attempts to place several speaking styles that have been studied in the literature (and others that have not yet, for contrast) into the three-dimensional space defined by the above axes. To give an example, a sportscast does not usually attain a high cultural level, is destined for a large unseen and therefore unfamiliar public, and, in the height of the action, gives less than maximal attention to intelligibility. Reading to a child finds a speaker who tries to be a good cultural vehicle before a very familiar listener trying to be well-understood.

Fine distinctions might be made although it is not evident that corresponding acoustic and perceptual evidence could be found to differentiate them. For example, speaking clearly for the hard-of-hearing, for attention-getting, for repetition and for a foreigner may not be different from one another, whereas they will differ from speaking clearly in noise (Lombard speech). As may be seen above therefore, speaking styles are not necessarily discrete entities with a clear difference between a given style and its "nearest neighbour".

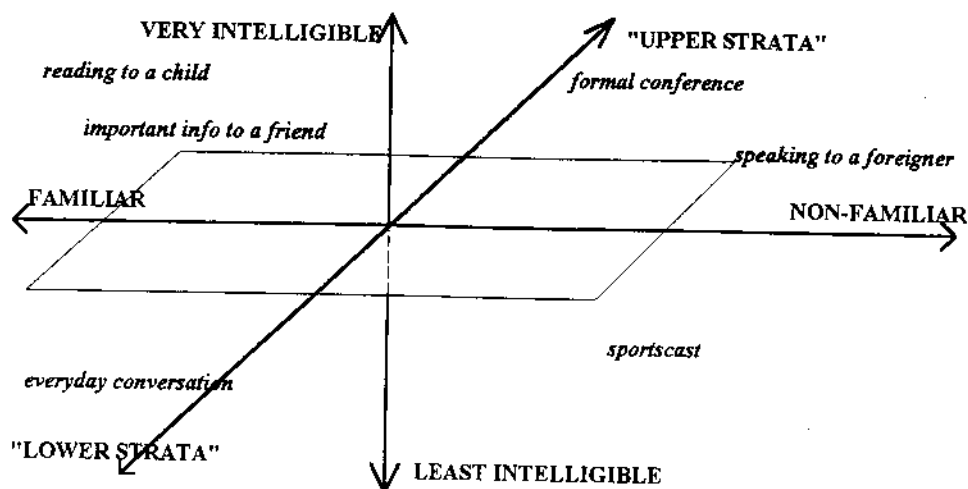


Figure 1. A three-dimensional style space

SPEAKING STYLE TERMINOLOGY IN THE LITERATURE

"Speaking styles" is a term that has been found to define many speech variations. A census of over 150 documents in the literature [EM92] finds the following terms, which we have divided into four categories:

A. Voice qualities: *breathy / creaky / lax / whispery modal / tense*. This seems to be an area in itself, voice qualities [LV91], contributing as a subelement to the impression of style change.

B. Speaking rate: *fast / very fast / slow*. This area also seems a subelement that aids in giving an impression of style change, but not a style per se.

C. Dimensions of speaking styles: *careful / clear / formal / citation / casual / less formal / conversational/free / spontaneous / connected / reiteration / repeated speech / scripted / unscripted / normal / reading / laboratory*. This area includes some of the denominations of our axes (or synonyms of them). Other terms here could be represented by a three-dimensional area in Figure 1, grouping several styles, such as "reading" for "reading to a child", "reading to a blind person", and "reading from a computer screen". Another of these terms is the ubiquitous "spontaneous".

D. Specific tasks and styles: *newscasts / sports / concerts / professional / interview*.

SOME SCHOOLS OF RESEARCH ON SPEAKING STYLES

The present era of research on speaking styles may be said to have begun with L. Shockey's thesis [SH73], which laid

the first foundations of seriously examining acoustico-phonetic evidence. Some other groups working on the subject:

Examining casual and clear speaking styles to determine what people do to make themselves better understood in order to ameliorate aids to the hard of hearing: Braida, Picheny, Uchanski, Chen, et al.

Style as a constantly changing entity, changing according to the evolution of the dialogue, examining the relation between duration and feature reduction: Lindblom, Engstrand, Krull, Duez, et al.

Focus words are not modified in the same way as non-focus words in style change: Koopmans van Beinum et al.

Individuals go about expressing the same perceived result using different factors: Eskénazi, Lacheret, Péan.

Direct experimentation in speech synthesis of the manipulation of various factors to obtain an impression of style change: Grandström, Carlson, Nord, Bladon.

Exploring the difference in speaking styles to try to find a way of obtaining data in a more cost-effective, but realistic way: Silverman et al.

FACTORS INDICATING STYLE CHANGE

Before estimating the factors linked to style change, it is necessary to prove that intended changes in speaking styles are actually perceived as such. [UC93], [RE91] and [PI86], for example, have tested the perceptual validity of their data, finding clear speech to be perceived as more intelligible by listeners, and in the case of the latter, by hearing-impaired listeners as well. [CH80] and [EI91] also found clear speech to be more intelligible, but only for a subset of their

speakers. [SO92], however, states the opposite, finding that all speakers are equally as intelligible in a given style. The reason for this apparent discrepancy may reside in the speech elicitation tactics used (see data collection below). [UC93], when eliciting a repetition of casual speech, finds it more intelligible than the first casual signal, but still less intelligible than clear speech.

The literature on speaking styles has touched on many diverse aspects, as can be seen in Figure 2.

| |
|---|
| <p>Acoustico-phonetic:</p> <p><u>Consonants</u>: flaps, stop releases, glottal stop, realisation of "th", burst frequency, alveolar stops, frication, non-syllabic consonants</p> <p><u>Formants</u>: transition rates, F2 onset, formants of diphthongs, formant frequency values</p> <p><u>Other vowel</u>: vowel reduction, VOT</p> <p>Prosodic:</p> <p><u>Duration</u>: syllable duration, plosive duration, vowel duration, silence duration, duration, speaking rate</p> <p><u>F0</u>: F0 range, median F0, maximum F0, F0 shape, F0 movement & intonation</p> <p><u>Other prosody</u>: amplitude, rhythm, pauses, spectral maxima</p> <p>Upper level:</p> <p><u>Phonology</u>: phoneme insertion, /R/ elision, consonant deletion, assimilation & elision, phonological reduction</p> <p><u>Word level and above</u>: word class & stress, pronouncing as reading, word length, verb form & sentence structure</p> <p><u>Focus words & content words</u>: content vs. non-content, focus vs. non-focus</p> <p><u>Redundancy & Intelligibility</u>: redundancy, intelligibility</p> |
|---|

Figure 2. Types of evidence of style change explored in the literature (according to the authors' own terminology)

A first glance at the evidence shows that there is no one factor alone that indicates style change, as has been noted by [RE91] and [EI91]. A closer glance shows that many studies have been based on little evidence - few speakers or small amounts of speech. We will try to relate only those factors which have statistical support, or we will indicate otherwise, if the factor bears mention.

Another remark concerns the conclusions that have been found. For a majority of the evidence that was examined in the making of this article, authors found inconclusive evidence for certain factors that they studied when they grouped their data over all speakers. They did, however, have statistically valid evidence for some speakers. This speaker-specific aspect will be mentioned below. It tends to lend credence to the hypothesis that different speakers use different strategies to obtain the same perceived result [EM92].

The results below come from data in many different languages. No language-related conflicts in the results have been noted, and the language studied is only mentioned where it sheds light on the pertinence of the results.

Articulation Factors

The attention to the articulation of speech has been studied, mostly to compare casual to clear speech, or "spontaneous" (usually casual, for example a conversation with a friend) to read speech. Attention to articulation is defined here to be the degree of attainment of articulatory targets (such as a given formant frequency or stop releases).

Consonants

The evidence concerning consonants generally agrees that articulatory targets are reached much more often in clear, or read speech than in casual.

For stops, [PI86] found that bursts were deleted in casual speech, but not in corresponding clear speech. [CH80] found that there was less of a scatter of frequency values for bursts in clear speech than in "conversational" (roughly casual) speech. Finally, [BW90] suggested that alveolar stops assimilate to labials or velars (according to context) although no statistical validation was given.

Fricatives have been studied from the point of view of the quantity of frication (noise) that is present. The realisation of /th/ in English, according to [LA64], is a fricative, an affricate, or a lenis stop, according to class strata as the speech gets more and more casual. [BN87] and reports in [GR92] describe reducing the amount of frication in synthesised speech to aid in obtaining a perceived change from formal (less frication) to casual (more) speech. And [PI86] has found a shift in the spectral maximum of the frication noise of /s/ toward a higher value in clear speech.

Vowels

transition rates

Transition rates, accounting for the movement of the formants at the onset and the offset of a vowel, reflect the coarticulation of the vowel with its neighbours and offer an indication of whether articulatory targets for the vowel itself may be attained (as we will see in the section on formant values below). Authors relate this to the casualness of speech.

[PI86] finds greater (longer) transition rates in clear speech, as does [CH80]. And the same findings are expressed by [KU89] and [DU92] in terms of more CV coarticulation (in terms of locus-nucleus values for the latter) in spontaneous (again casual) speech. It should be noted that [CH80] did not find this true for all speakers.

[LB90], studying the formants of diphthongs, found that the values of the formants of /aI/ in English only attain target values when duration is long, thus not usually in casual speech.

formant values

Much work has concerned formant (especially F1/F2) values. Some has centred around the size of the cluster of these values (variability), and some has concerned target attainment, that is, whether these values tend toward the extremes of the "vowel triangle" or whether targets are not as well attained, with more "central" formant values.

[CH80], [PI85], [HO92], [VB91], and [KU89] all find formant undershoot in casual speech, especially for dentals, for the latter. [CH80] finds also that a plot of F1/F2 values shows a much smaller cluster for clear speech than for casual. More specifically, [KO92] finds that F1 is more variable than F2, and [PI86] does not find smaller clusters of F1/F2 for clear speech when plotting only tense vowels. [GA78] has shown that stress, not duration, affects target attainment in English. [EO89] also showed that short durations do not always lead to formant undershoot in Swedish, therefore implying that vowel reduction and speaking rate are mainly independent of one another.

Concerning speaker specifics, [CH80] found statistically significant results including F3 (variability) for some speakers, but not for all, and [GA78] did not have significant results on F1/F2 for a part of his speaker population.

VOT

[CH80] has found that the difference in VOT (voice onset time) between voiced and voiceless consonants (usually shorter for the former) is greater in clear speech, adding another argument to the defence of increased articulatory effort in clear speech.

Prosody

duration

Changes in duration have been studied for speech that was simply faster or slower, but also for casual as opposed to clear, and casual as opposed to read speech. Many aspects have been studied since, as [PI89] and [CH80] noted, changes in duration are not uniform over the whole speech signal. And, from the point of perception, [RE86] found that the one element that most successfully allows a listener to distinguish casual from read speech is "phonatory frequency". [UC93] has more precisely found that non-uniform changes in duration make speech more comprehensible than uniform ones, be it for slowing down, or for speeding up speech.

Clear speech has, in one study, [PI86], been shown to be twice as long as casual speech and speech in noise has also been found to be characterised by longer duration [CL87]. Only one study [WC89] found that duration increases from fast to slow speech were statistically significant for only a part of the speaker population (and this especially concerning foot structure in stressed vs. unstressed positions).

In as far as consonants are concerned, [PI86] found that overall plosive duration increased in clear speech while [CH80] went further, measuring the silence interval separately for labials, velars, and dentals, with the same general findings, the relation between the durations for the different places of articulation being speaker-specific. [DU92] found that for read as opposed to casual speech, dentals and labials were significantly longer in French. [UC93] also noted that unvoiced fricatives and semivowels increased more than vowels and voiced fricatives in clear speech.

For vowel duration, [CH80] found that the duration of clear vowels increases proportionately (by type of vowel) to the duration of casual ones. [GA78], when examining fast, "normal", and slow speech, attributed the major change in duration to the change in the duration of the vowels. This might seem slightly in contradiction to other information at a first glance, but if we take into consideration that this was an elicited change in the rate of reading rather than in a speaking style, the other elements we have seen which speakers use to make their speech clearer would not have been used by these subjects.

It is also interesting to note the work of [CH80] and [EI91] on the placement of pauses (and this the alteration of the total duration) around important content words in clear speech.

The literature therefore has so far noted that when speakers lengthen their speech when trying to speak more clearly or in reading, they lengthen the stable parts of the signal more than the transitional parts (but note the work on vowel transitions).

speaking rate

A rapid glance at overall change in speaking rate alone concerning perception shows that faster speech is not less intelligible to adults [SO92] although when the listeners are children, the opposite is true [BE73].

F0

Speaking styles have been compared to determine whether changes in fundamental frequency, F0, may be significant. Authors have looked at maximum F0, median F0, F0 range and F0 movement.

[PI86] noted slightly higher F0 maxima in clear rather than casual speech. Concerning the median value of F0, refined results in [KO91], [KV91] and [KO92] find median F0 to be higher in read speech. [CL87] found median F0 to be higher in noise.

Concerning F0 range, [PI86] found it to be greater in clear speech than in casual speech while [ES92] found it to be greater in read speech than in either clear or casual speech, for a subset of speakers. [AY91] found no difference between read and casual speech, but this may be due to the undefined, thus rather neutral, reading style here.

F0 movement was found by [AY91] to have a wider variety of contour types in casual than in read speech. [SI92] and [SB92] show that city names are pronounced with a rising (polite) intonation in American English in casual speech whereas they have a falling intonation in read speech. [BL92] found that F0 excursions in the accented part of read speech are larger and steeper than in the accented part of casual speech.

amplitude

Little evidence has been found on the subject of the intensity of the speech signal. Although [CL87] has found higher long term spectral energy in Lombard speech, work concerning clear speech [PI86] has only found a subpopulation of speakers who speak louder when trying to be clearer. This has led [PI86] and [HG92] to note that many prosodic elements interact in speaking styles and that no one alone, especially amplitude, may be used to achieve the desired style.

In the domain of amplitude, note must be given to the synthesis of strong vs. soft speech using amplitude variations on the GLOVE synthesiser [GN92].

pauses

The study of the use of pauses in style change has shown some very significant results. As mentioned above, speech is lengthened in clear speech by the insertion of pauses at boundaries of high-content words [CB90]. [HI91] noted that there are more pauses within a tone unit in British English in casual speech than in reading. [SI92], [SB92], and [GO68] have all noted that there are many more ungrammatical pauses in casual speech than in read speech. These pauses are often longer and more variable in length, and [SB92] found no filled pauses in read speech. It should, however, be noted that ([GO68]) pause lengths differ greatly from one speaker to another and make up a variable (10 to 60%) portion of the total duration of speech.

Phonology

Most studies on the phonological level concern the various reduced phonetic forms which appear as speech style

becomes casual. In general, as [PI85] observed, there is increasing phonological variability in casual speech. [GM89], observing fast speech, noted the increased amount of elisions. Yet it may be noted that several authors have warned that increased speaking rate does not have a direct link with, for example, reduced vowels. Rather, it is felt that other prosodic elements also have a role in reduction. [BW90] noted the possibility of simultaneous deletion and assimilation in casual speech.

More specifically, [PI86] noted that word-final voiced consonants are often followed by a schwa in clear speech and a schwa may be inserted in a voiced fricative-semivowel cluster in the same style. [LA64] noted increased /R/ elision as speech became less formal. [EL91] noted the pronunciation of the final /l/ in words like "nombri" in read French for some speakers. And, [BW90] noted the reduction of the alveolar gesture as speech became less casual.

In speech synthesis, [BN87] tried to change synthetic speech along a formal-to-casual axis by using increasingly reduced forms of common function words such as "can".

Remarks

In conclusion to this section, we might try to plot some of the significant factors from the literature into our style space (Figure 3).

AUTOMATIC SPEECH PROCESSING

As mentioned in passing above, there have been efforts, notably at KTH, to synthesise changes in speaking styles using either phonologically reduction to go from formal to casual speech, or amplitude to go from soft to loud speech.

In speech recognition, NYNEX has examined casual and read speaking styles to see if they could obtain natural speech at less cost, eliciting naturalness from reading - this has not yet given satisfactory results.

DATABASES

It is difficult to obtain significant quantities of natural speech, especially when researchers want to compare different speaking styles for the same linguistic content, or the same phonetic context. Many of the early work consisted of hiding a microphone, or not controlling production in any way, and then comparing the results to a screen- or paper-read transcription. The conditions were totally different, and the results were, of course, quite different, even if nothing could be generalised based on them. Elicitation techniques have evolved greatly in the past few years. Following are a few examples.

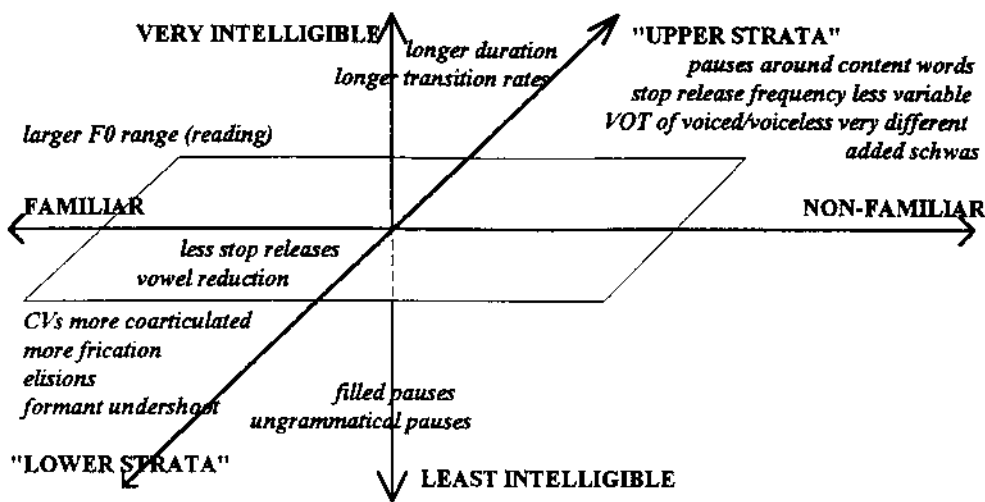


Figure 3. Some pertinent factors plotted in the aforedefined style space

The main concern in data collecting has been naturalness. The work that has been carried out that seeks comparable linguistic content in more than one speaking style can be viewed as being on an axis going from least to most natural. The three points along the axis, corresponding to types of data collection are: scenario, game, real task.

Scenario

The speaker in this type of elicitation is to forget the microphone before him by involving himself in a role to be played. This task, since it is fairly far from the speaker's everyday concerns, generally meets with mixed success and only data that has been style-verified is usable. [EH90] has speakers pretend that they are registering for a conference, and [EI91] had speakers call the laboratory to ask for three-month student project information.

Games

Involving a speaker in playing a game lets him forget the microphone and furnishes really natural speech. The data collected here, however, is task-dependant, that is, results cannot easily be extrapolated to other tasks. The linguistic content is, however, closely controlled, and the speaking styles chosen may be comparable instead of the old conversation-with-a-friend/screen-reading paradigm.

[MC89] and [BB90] have set up a database based on the description of a path in a map with phonological, lexical, syntactic and semantic control of speaker output. The paradigm may compare familiarity since the speaker population was chosen to be of different degrees of this quality. [SZ92] sends speakers through Levelt's spatial gridlike network, describing geometrical objects, with strong content control.

[GO68] had subjects describe a series of pictures in a cartoon story form six times. The work by [VP93] has speakers play a game of seven errors (in comparing two

pictures where the "errors" are phonological contexts where variation is likely to occur), and changes styles from casual to clear to read, all for the same context. This is the only data to our knowledge that enables the study of more than one style change.

real tasks

To our knowledge, only one collection of data has taken place where speakers were performing an everyday task in an everyday environment. This is the work described in [SI92] and [SB92], which also has the largest speaker population. The speakers were calling directory assistance; their speech was later transcribed and reread with instructions to be as natural as possible.

CONCLUSION

Many factors have been studied in trying to characterise the change in speaking styles, but little automatic speech processing work has been done. This may be for the following reasons:

- 1) The concept of speaking styles has to present been loosely defined with little theoretical basis. Such aspects as the interaction of the speaker and his environment has not yet been dealt with, for example.
- 2) Contractual work on the subject has only taken off with the work of the DARPA community on increasingly natural speech and of the synthesis community on increasingly natural synthetic voices.
- 3) The data has been sparse and has not compared styles which have any practical relation to one another, and each factor studied, with few exceptions, has been studied as a separate entity, not yet related to the others.

The beginning of this decade had seen new, promising work in this domain which should make research on speaking

styles soar. Another aspect which has been practically ignored so far is the relation of speaking style and cognitive workload which [GO68] and [LI93] have examined, an element which contributes to the way speakers attack style variations.

ACKNOWLEDGEMENTS

We beg the indulgence of authors of work that may have been overlooked above, and request that they send references to: max@limsi.fr. Also, please excuse the lack of reference to the name of the first author in many of the references within the article - this is solely due to lack of space. The articles cited here make up only a part of the literature in the domain and the reader is encouraged to obtain [EM92] for a fuller list. In this article we have not mentioned articulatory work such as that by Shockey and Farnetani, and the database work by Howell. We wish to thank J.S. Liénard, and V. Péan for their comments on the scientific content of the article, and C. Leleu, S. Faye, and G. Lacoste for their help in reference research. This article has been written in direct relation with the ESPRIT VOX Working Group no. 6298 activities.

REFERENCES

[AY91] Ayers, G., 1991, "A comparative intonational analysis of spontaneous and read speech", *Journal Acoust. Soc. Am.*, vol. 79 supp.1, p. S26.
 [BB90] Bard, E.G., Anderson, A.H., Garrod, S., Isard, S.D., Miller, J.E., McAllister, J., Bennett, C., Kostko, J., Weinert, R., Boyle, L., 1990, "A vertical study of dialogue", Report P05 of HCRC, University of Edinburgh, Edinburgh.
 [BE73] Berry, M., Erickson, R., 1973, "Speaking rate: effects on children's comprehension of normal speech", *JSHR*, no. 16, p. 367 - 374.
 [BL92] Blaauw, E., 1992, "Phonetic differences between read and spontaneous speech", *Proc. ICSLP Conference, Banff*, p. 751 - 754.
 [BN87] Bladon, A., Carlson, R., Granström, B., Hunnicutt, S., Karlsson, L., 1987, "A text-to-speech system for British English, and issues of dialect and style", *European Conference on Speech Technology, Edinburgh*, p. 55 - 58.
 [BW90] Browman, C., Goldstein, L., 1990, "Tiers in articulatory phonology with some implications for casual speech", in *Papers in Laboratory Phonology I*, ed. Kingston, J. and Beckman, M., Cambridge University Press, p. 341-376.
 [CB90] Cutler, A., Butterfield, S., 1990, "Durational cues to word boundaries in clear speech", *Speech Communication*, vol. 9, p. 485 - 495.
 [CH80] Chen, F., 1980, "Acoustic Characteristics and Intelligibility of Clear and Conversational Speech at the Segmental Level", Master's thesis, Massachusetts Institute of Technology - Dept. of Electrical Engineering, dir. V. Zue.
 [CL87] Clark, J.E., Lubker, J.F., Hunnicutt, S., 1987, "Some preliminary evidence for phonetic adjustment strategies in communication difficulty", in *Language Topics*, J. Benjamins.
 [DU92] Duez, D., 1992, "Second formant locus-nucleus patterns: an investigation of spontaneous French speech", *Speech Communication*, vol. 11, no. 4 - 5, p. 417 - 428.
 [EH90] Ehara, T., 1990, "Token count statistics extracted from the ATR dialogue database", ATR Technical Report TR-1-0157.
 [EI91] Eskénazi, M., Isard, A., 1991, "Characterising the change from casual to careful style in spontaneous speech", *JASA*, vol. 90 no. 4 pt. 2, p. 2363 - 2364.
 [EL91] Eskénazi, M., Lacheret-Dujour, A., 1991, "Exploration of individual strategies in continuous speech", *Speech Communication* 10, p.249 - 264.
 [EM92] Eskénazi, M., 1992, "A bibliography on speaking styles and related subjects: part 1", *Notes et Documents du LIMSI*, no. 92-14.
 [EO89] Engstrand, O., 1989, "F0 correlates of tonal word accents in spontaneous speech: range and systematicity of variation", *PERILUS*, University of Stockholm, no. X, p. 1 - 12.

[ES92] Eskénazi, M., 1992, "Changing speech styles: speakers' strategies in read speech and careful and casual spontaneous speech", *Proc. ICSLP Conference, Banff*, p. 755 - 758.
 [GA78] Gay, T., 1978, "Effect of speaking rate on vowel formant movements", *JASA*, vol. 63 no. 1, p. 223 - 230.
 [GM89] Gimson, A.G., 1989, *An Introduction to the Pronunciation of English*, (4th edition, S. Ramsaran, ed.), Edward Arnold, London, p. 307 - 311.
 [GN92] Granström, B., Nord, L., 1992, "Neglected dimensions in speech synthesis", *Speech Communication*, vol. 11, no. 4 - 5, p. 459 - 462.
 [GO68] Goldman-Eisler, F., 1968, *Psycholinguistics: experiments in spontaneous speech*, Academic Press, London.
 [GR92] Granström, B., 1992, "The use of speech synthesis in exploring different speaking styles", *Speech Communication*, vol. 11, no. 4 - 5, p. 347 - 356.
 [HG92] Hirschberg, J., Grosz, B., 1992, "Intonational features of local and global discourse structure", *Proc. Fifth DARPA Workshop on Speech and Natural Language*, Harriman N.Y., February 1992.
 [HI91] Hieronymus, J., Williams, B., 1991, "A comparison of the prosody in read speech and directed monologue in British English", *Proc. ESCA ETRW Phonetics and Phonology of Speaking Styles, Barcelona*, p. 32-1 - 32-5.
 [HO92] Harnegnies, B., Poch-Olive, D., 1992, "A study of style-induced vowel variability: laboratory versus spontaneous speech in Spanish", *Speech Communication*, vol. 11, no. 4 - 5, p. 429 - 438.
 [JK76] Jakobson, R., 1976, *Six leçons sur le son et le sens*, Editions de Minuit, Paris.
 [JO68] Joos, M., 1968, "The isolation of styles", in *Readings in the Sociology of Language*, ed. Fishman, J.A., Mouton, The Hague, p. 185 - 191.
 [KO91] Koopmans-van Beinum, F.J., 1991, "A peak and level model for focus words in read and spontaneous natural speech and in synthetic speech", *Proc. ESCA Eurospeech, Genoa*, p. 1459 - 1462.
 [KO92] Koopmans-van Beinum, F.J., 1992, "The role of focus words in natural and in synthetic continuous speech: acoustic aspects", *Speech Communication*, vol. 11 no. 4-5, p. 439 - 452.
 [KU89] Krull, D., 1989, "Second formant locus patterns and consonant-vowel coarticulation in spontaneous speech", *PERILUS*, University of Stockholm, no. X, p. 87 - 108.
 [KV91] Koopmans-van Beinum, F.J., 1991, "Spectro-temporal reduction and expansion in spontaneous speech and read text", *Proc. ESCA ETRW Phonetics and Phonology of Speaking Styles, Barcelona*, p. 36-1 - 36-5.
 [LA64] Labov, W., 1964, "Phonological correlates of social stratification", *American Anthropologist*, vol. 66, p. 164 - 176.
 [LA72] Labov, W., 1972, *Sociolinguistic Patterns*, University of Pennsylvania Press.
 [LB90] Lindblom, B., 1990, "Explaining phonetic variation: A sketch of the H and H theory", in *Speech Production and Speech Modelling*, ed. Hardcastle, W., Marchal, A., Kluwer, Dordrecht, p. 403 - 440.
 [LB91] Lindblom, B., 1991, "Speech transforms: on the extent, systematic nature and functional significance of phonetic variation", *Speech Communication*, vol. 11, no. 4 - 5, p. 357 - 368.
 [LI92] Listerri, J., 1992, "Speaking styles in speech research", *Proc. ELSNET/ESCA/SALT Workshop on Integrating Speech and Natural Language*, Dublin, Ireland.
 [LI93] Lively, S., Pisoni, D., Van Summers, W., Bernacki, R., 1993, "Effects of cognitive workload on speech production: Acoustic analyses and perceptual consequences", *Journal Acoust. Soc. Am.*, vol. 93 no. 5, p. 2962-2973.
 [LR84] Liénard, J.S., 1984, "Une approche globaliste de la variabilité acoustico-phonétique de la parole", *Proceedings, Journées d'Etude sur la Parole GALF*, Brussels.
 [LR90] Liénard, J.S., 1990, "Perception data variability and inductive inference", *Cognitiva* 90, Madrid.
 [LV91] Laver, J., 1991, *The Gift of Speech*, Edinburgh University Press, Edinburgh, Scotland.
 [MC89] McAllister, J., Sotillo, C., Bard, E., Anderson, A., 1989, "Using the map task to investigate variability in speech", *Occasional Paper*, University of Edinburgh, Dept. of Linguistics.
 [MU93] Murray, I., Arnott, J., 1993, "Toward the simulation of emotion in synthetic speech: a review of the literature on human vocal emotion", *J. Acoust. Soc. Am.*, vol. 93, no. 2, p. 1097-1108.
 [PI85] Picheny, M.A., Durlach, N., Braidia, L.D., 1985, "Speaking clearly for the hard of hearing I: Intelligibility differences between clear and conversational speech", *JSHR*, vol. 28, p. 96 - 103.
 [PI86] Picheny, M.A., Durlach, N., Braidia, L.D., 1986, "Speaking clearly for the hard of hearing II: acoustic characteristics of clear and conversational speech", *JSHR*, vol. 29, p. 434 - 446.

- [PI89] Picheny, M.A., Durlach, N.I., Braida, L.D., 1989, "Speaking clearly for the hard of hearing III: An attempt to determine the contribution of speaking rate to differences in intelligibility between clear and conversational speech", *JSHR*, vol. 32, p. 600 - 603.
- [RE86] Remez, R., Rubin, P., Nygaard, L., 1986, "On spontaneous speech and fluently spoken text: Production differences and perceptual distinctions", *Journal Acoust. Soc. Am.*, vol. 79 supp.1, p. S26.
- [RE91] Remez, R., Berns, S., Nutter, J., Lang, J., Davachi, L., Rubin, P., 1991, "On the perceptual differentiation of spontaneous and prepared speech", *Journal Acoust. Soc. Am.*, vol. 89 no. 4 pt.2, P. 2011-2012.
- [SB92] Silverman, K., Blaauw, E., Spitz, J., Pitrelli, J., 1992, "A prosodic comparison of spontaneous speech and read speech", *Proc. ICSLP Conference, Banff*, p. 1299 - 1302.
- [SH73] Shockey, L., 1973, "Phonetic and phonological properties of connected speech", Doctoral dissertation, Ohio State University, Dept. of Linguistics.
- [SI92] Silverman, K., Blaauw, E., Spitz, J., Pitrelli, J., 1992, "Towards using prosody in speech recognition/understanding systems: differences between read and spontaneous speech", *Proc. Fifth DARPA Workshop on Speech and Natural Language, Harriman N.Y.*, February 1992.
- [SO92] Sommers, M., Nygaard, L., Pisoni, D., 1992, "Stimulus variability and the perception of spoken words: effects of variation in speaking rate and overall amplitude", *Proc. ICSLP Conference, Banff*, p. 217 - 220.
- [SZ92] Swerts, M., Collier, R., 1992, "On the controlled elicitation of spontaneous speech", *Speech Communication*, vol. 11, no. 4 - 5, p. 463 - 468.
- [UC93] Uchanski, R. M., Braida, L. D., Reed, C.M., Choi, S., and Durlach, N. I. "Speaking Clearly for the Hard of Hearing IV: Further Studies of the Role of Speaking Rate" (1993) *JSHR* (to appear).
- [VB91] van Bergem, D., 1991, "The influence of sentence accent, word stress, and word class on the quality of vowels", *Proc. ESCA ETRW Phonetics and Phonology of Speaking Styles*, p. 10-1 - 10-5.
- [VP93] Péan, V., Williams, S., Eskénazi, M., 1993, "The design and recording of ICY, a corpus for the study of intraspeaker variability and the characterisation of speaking styles", *Proceedings of this conference*.
- [WC89] Campbell, W.N., 1989, "Timing Invariance in read speech", *Proc. ESCA ETRW Speaker Characterisation, Edinburgh*, p. 78 - 82.
- [ZW72] Zwicky, A., 1972, "On casual speech", in Papers from the Eighth Regional Meeting of the Chicago Linguistic Society, Chicago: Chicago Linguistic Society, p. 607 - 615.