Speech synthesis of dialectal variants as a method for research on prosody

Beat Siebenhaar, Martin Forst, Eric Keller
Laboratoire d'Analyse Informatique de la Parole, Faculté des Lettres, Université de Lausanne, CH-1015 Schweiz

Speech synthesis has reached quality levels that henceforth permit its use in testing linguistic hypotheses (Keller and Zellner Keller 2000). We wish to present a Swiss National Fund study on dialectal prosody which initial goal is to build a complete synthesis of two Swiss German dialects. On the one hand, the two synthesis systems will let us compare globally the prosody of the two dialects. On the other hand, they will let us change the utterances in a consistent fashion to investigate dialectal variation of prosody.

The sound examples for this article are available at http://www.unil.ch/imm/docs/LAIP/wav.files/MethodsXISnd.zip. The sound files are named 'Ex_01_Title', 'Ex_02_Title', etc. and are referred to in this manner in the text. The transcription of the examples follows the principles of Dieth (1986) first published in 1938.

1. Problems of Prosody Modification

In linguistic research one often confronts subjects with speech signals in order to rate them for linguistic adequacy or to obtain information about the local or social situation of the speaker (for Swiss German dialects cf. Werlen 1980, Werlen 1985, Hengartner 1995, Hofer 1997; for Alemannic cf. Gilles and Schrambke 2000). Quite often, matched guise techniques are used. In this procedure, developed by W. E. Lambert, a bilingual/bidialectal person is recorded in her/his different varieties. The test persons are then asked to rate some of these recordings, not knowing that two of them are from the same speaker. When the ratings of the two recordings of the same speaker are compared, it is supposed that these ratings are not dependent on the voice characteristics of the speaker, but only on the linguistic differences of the recordings. For research on prosody, it is very hard to obtain such stimuli, since prosodic parameters are difficult to control. Therefore, studies often rely on manipulating the original speech signals, which can be done relatively easily with modern phonetic instruments. Thus, while manipulating a speech signal is less of a technical problem nowadays then it was until recently, it remains a linguistic problem. We can perform almost any manipulation we like, but we still do not know what is allowed by the prosodic grammar of a given language.

For example, the speed at which f0 can rise or fall is constrained by physiological limitations of the glottis. When changing segment durations, one might shorten a given portion of speech so much that the corresponding f0 contour ends up too steep to be considered natural. Furthermore, there are not only physiological limitations, but even more important constraints given by the grammar limiting the steepness of f0 contours at certain places. In addition, one has to take into account the complex interactions and interrelations between duration and f0, and simple linear changes of segment duration or of a part of an utterance do not give realistic results.

We will illustrate this problem of changing the prosodic parameters with some examples of manually modified sound files. The manipulation tool used for this task is Praat 4.0[1]. The original sound files are comparison texts for Swiss German dialects recorded in the 1940's to 1960's (Phonogrammarchiv der Universität Zurich 2000). These texts share most of the occurring lexical elements, so that in most cases where there are no syntactical or lexical differences, it is relatively easy to compare the data and to modify them in parallel. Let us now look at the recordings of the Bernese and Zurich dialects. (Ex_01_Zuerich.wav, Ex_02_Bern.wav). The text is 'I wish you too a good new year'. The figures show the f0 contour and the words in the time domain.

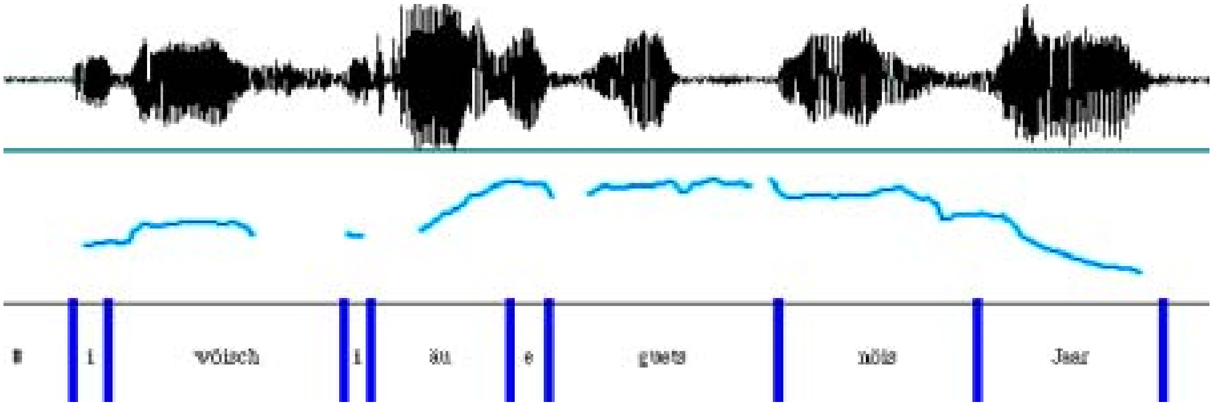

Figure 1: Zurich original: *I wöisch i äu e guets nöis Jaar.* 'I wish you too a good new year'. (Ex_01_Zuerich.wav)

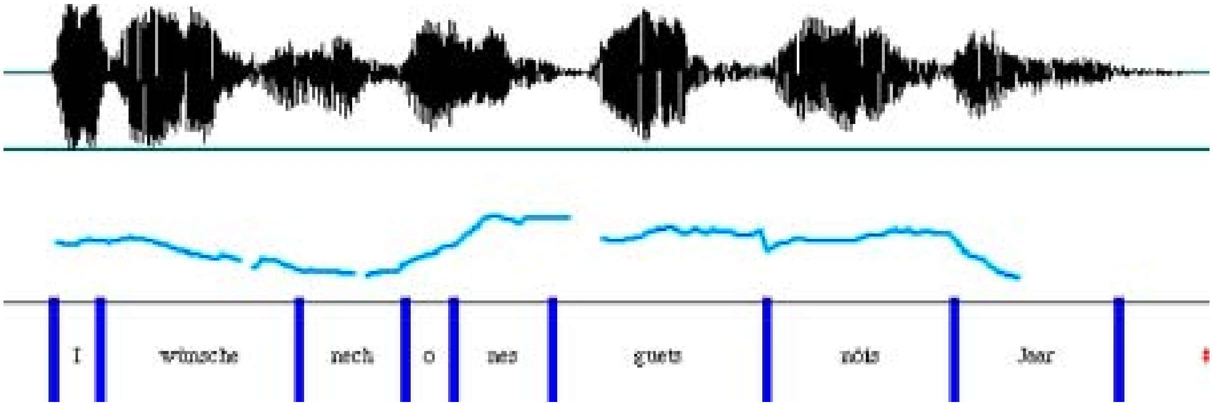

Figure 2: Bernese original: *I wünsche nech o nes guets nöis Jaar.* 'I wish you too a good new year' (Ex_02_Bern.wav)

We now wish to to transplant the prosodics from one dialect to the other one, while maintaining the segmental information. The time slots are given by the length of the words, which are all monosyllabic except for the verb, which is bisyllabic in the Bernese dialect. This copy prosody illustrates the problem of interrelation of timing and f0. In Figure 3 the Bernese word durations have been transferred to the Zurich sound file. (Ex_03_Zuerich_BETim.wav). This results in an f0 that sounds very unnatural, mainly because the f0 rise in *äu* is too steep.

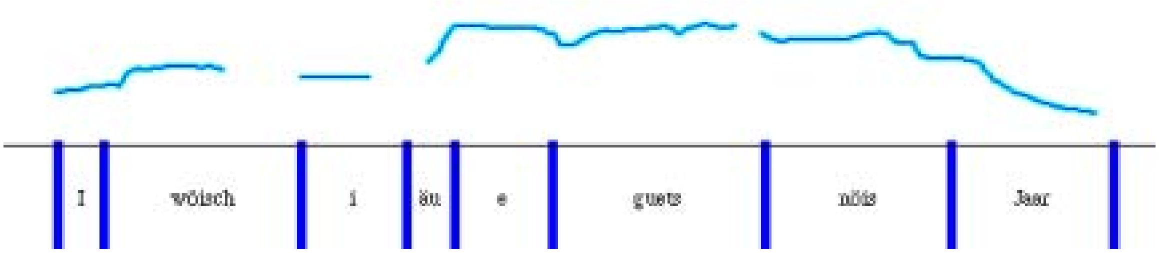

Figure 3: Zurich sound with Bernese Timing (Ex_03_Zuerich_BETim.wav)

It thus is obvious that the changes in timing also have an influence on the f0 contour. The same is true if one only changes the f0 information (Ex_04_Zuerich_BEInt.wav, Figure 4) without taking the timing information into account. Here the rise in the article *e* occurs too fast, giving it a unusual and unintended prominence.

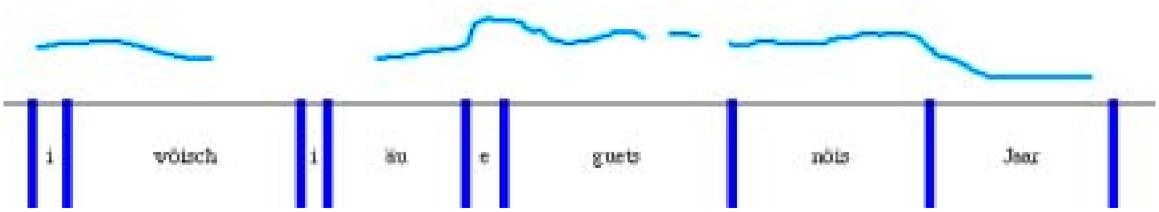

Figure 4: Zurich sound with Bernese intonation (Ex_04_Zuerich_BEInt.wav)

So if we want to change the speed of a part of a given utterance, we have to change its intonation curve *and* its timing information. Figure 5 (Ex_05_Zuerich_BETimInt.wav) show the transplantation of the whole prosodic information from one sound file to the other.

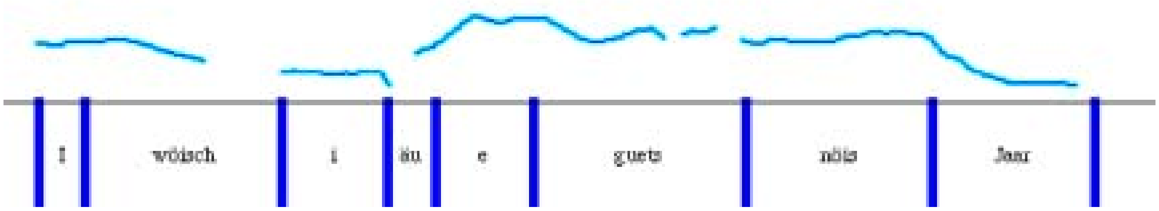

Figure 5: Zurich sound with Bernese timing and intonation (Ex_05_Zuerich_BETimInt.wav)

The same problems occur when the Zurich prosodic information is transferred to the Bernese text. Figure 6 shows the transplantation of timing and here especially the very fast f0 rise in *nes* while the f0 contour on the *o*, with originally was focussed, is too flat.

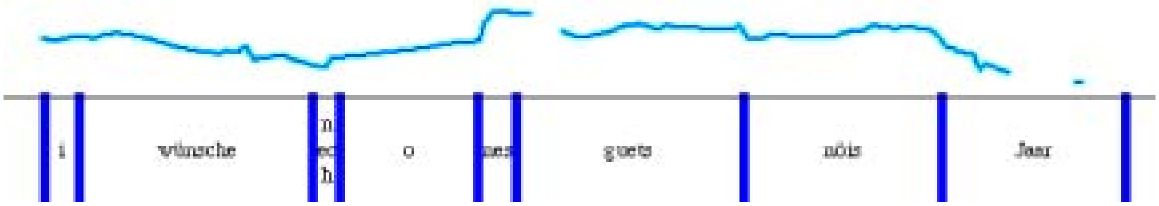

Figure 6: Bernese sound with Zurich timing (Ex_06_Bern_ZHTim.wav)

Figure 7 shows the transplantation of Zurich intonation curve to the Bernese segments and its timing information.

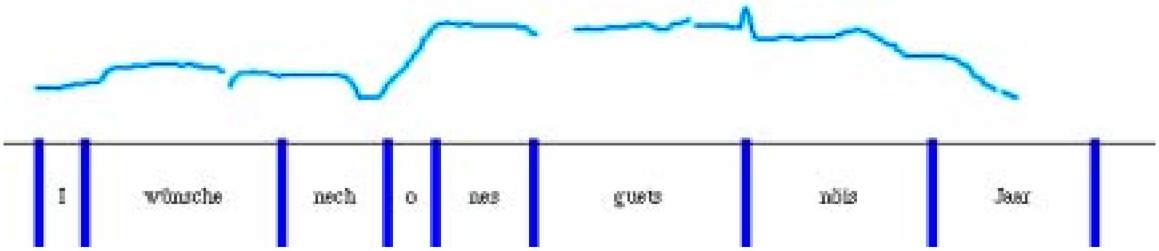

Figure 7: Bernese sound with Zurich intonation (Ex_07_Bern_ZHInt.wav)

In Figure 8 both timing and intonation of the Zurich sound file are applied to the Bernese segments.

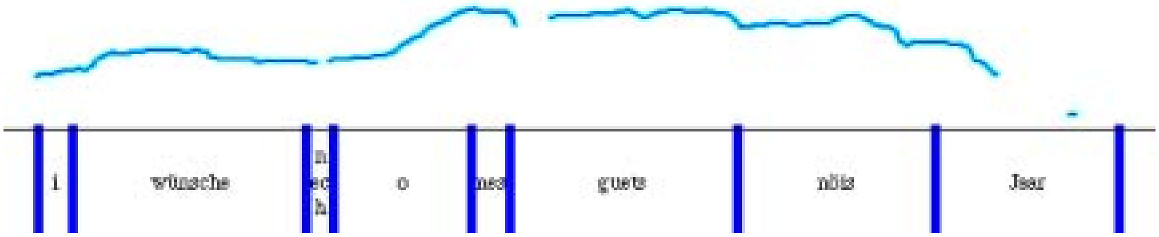

Figure 8: Bernese sound with Zurich timing and intonation (Ex_08_Bern_ZHTimInt.wav)

These examples demonstrate some of the major problems of prosody modification. Even in a prosody transplantation of almost identical texts, the result is not really as expected; in our example, the first half of the signal does not meet our expectations. The two dialects have a different number of segments in certain syllables - *i* vs. *nech* for the dative personal pronoun and *e* vs. *nes* for the neutral accusative article - and they have a different number of syllables in the verb - *woisch* vs. *wunsche*. At these places, the prosodics do not match the segmental material and the transplantation of the prosody does not really fit. For a better fit of the curves, we would have to respect the various segmental aspects. It is almost impossible to make all these changes manually in a consistent manner if one does not have truly matching signals; and even small differences in the segmental sequence affect the prosodic characteristics of the signal. Moreover, as soon as one also wants to change syntactic, morphologic or phonetic aspects or the sound quality and its characteristics, this task becomes completely impossible.

Our solution to this problem is to use speech synthesis as a linguistic research tool. In the following we will outline this approach, its specific needs, advantages and disadvantages.

2. Speech Synthesis as a Tool for Linguistic Research

The traditional scientific procedure to obtain information consists in analyzing data and explaining the findings concerning a certain phenomenon. The advantage as well as the disadvantage of this procedure is that apparently non-relevant phenomena are excluded.

For instance, in research on intonation work is focused on f0 contours. Specific contours are selected from a corpus, they are described and compared to other contours in similar and different environments. They are classified into highs and lows, early and late highs, boundary tones. After some hard work, a sophisticated description of a specified aspect of intonation is obtained, but quite often the repercussions that rhythmic or segmental phenomena may have on such f0 contours are neglected. Similar remarks can be made for analysis in any research field.

Using synthesis as a scientific method allows us, indeed forces us, to model interactions between the different subsystems that are typically set aside in an exclusively analytic procedure. For the investigation of prosody this means that the design of a synthesis system has to take into account all aspects of prosody and has to reflect the interrelation between the different linguistic levels, the very aspects that were just shown to pose problems in prosody transplantation. Therefore a speech synthesis system can be used to test the interplay of segmental and suprasegmental information, of phrasing, timing and intonation.

Some examples with obvious errors in specific aspects of prosody may illustrate this fact. The examples are synthesized with our TTS-System for standard German, which permits us to control different parameters:

1. Imagine how difficult it would be to understand a speech signal where phrase boundaries are placed in an arbitrary fashion. *Stellen Sie# sich vor, wie# schwer es ware ein Sprachsignal zu# verstehen, wo Phrasengrenzen# vollkommen# willkurlich# gesetzt werden.* (Ex_09_Falsche_Phrasengrenzen.wav)

2. Imagine a speech signal where the underlying timing model does not take phrase boundaries into account. You hear that the sounds before pauses are too short because, among other things, the so-called final lengthening is not respected. Pauses are therefore badly perceived; the utterance sounds choppy. *Stellen Sie sich ein Sprachsignal vor, in dem das zugrundeliegende Modell fur die zeitliche Steuerung keine Phrasengrenzen berucksichtigt. Sie horen, dass die Laute vor den Pausen zu kurz sind, weil unter anderem keine Langung am Phrasenende eingebaut ist. Pausen werden so relativ schlecht wahrgenommen; die Sprache wirkt abgehackt.* (Ex_10_Ohne_Phrasentiming.wav)

These examples show that a speech synthesis system must implement adequate models for all aspects of prosody. A deficient model for one aspect can obscure the adequacy of all other aspects in the resulting speech output.

With this background, we are ready to approach dialectal prosody. This means that instead of carrying out an analysis of some specific aspect of the prosody of a given dialect, we intend to build a complete model of prosody generation. This holistic approach has certainly some drawbacks[2] – we do not deny that –, but they differ from those of a purely analytic approach, so that we might obtain complementary information about prosody. While the analysis of particular aspects allows us to obtain detailed information about the selected phenomena, the synthesis approach forces us to consider all aspects of speech, and thus may point out where we still lack sufficient information for an implementation. This is a result that often leads to further analysis. In this sense, speech synthesis is a diagnostic tool. On the other hand, we cannot stick to details without having clarified the global aspects. So synthesis forces us to elaborate global aspects first, which can then be refined step by step.

3. Architecture and History of LAIPTTS

As the design of our dialectal synthesis system will follow the 'LAIP tradition' of speech synthesis concerning major theoretical decisions, we will briefly present the basics of the psycholinguistically and statistically motivated models of phrasing, timing, intonation and their interplay that Keller and Zellner Keller elaborated for French (Keller and Zellner, 1998). They are at least in part implemented in the French speech synthesis system LAIPTTS_F, and their adaptation to standard German has lead to the development of LAIPTTS_D[3]. Since this first adaptation was reasonably successful we now employ speech synthesis as a prosodic research tool. We are now in the process of building a Bernese and a Zurich German synthesis system according to the same principles.

The input to our prosodic module is the phonetic chain, annotated for word and syllable boundaries as well as the grammatical or lexical status of words. This phonetic chain is split up into prosodic phrases. Augmented with information concerning phrase boundaries, it is the basis for the calculation of the duration of the single segments. As a final step, the f0 contour is calculated on the basis of the phonetic material, the corresponding durations and phrase boundaries.

Thus it is assumed in our systems that the phrasing component precedes the temporal calculation component, and that the temporal calculation component precedes the intonation component. This conception is based on the fact that any human action process is necessarily embedded in a temporal structure. Furthermore it has been shown that the durational domain is subject to more rigid constraints than the f0 domain. For example, it was shown by Keller (1994, based on data by Caelen-Haumont, 1991) that within a given syllable duration correlates much more between speakers than f0.

Because of this linear succession of speech generation modules, our speech synthesis models can represent influences of phrasing on segment duration, and influences of timing on intonation. But there is no influence from intonation to phrasing or segment duration in this architecture. The model is therefore a linear model without recursivity. The different components of the model will now be explained in detail.

3.1. Phrasing

Speaking is constrained by human cognitive and physiological abilities. For this reason, human beings structure utterances into intelligible parts that can be handled by these abilities. The most evident structuring elements in speech are pauses, both silent ones, where there is a complete absence of sound, and pauses that are filled by hesitation or respiration sounds. However, prosodic phrases are not necessarily separated by pauses; they can also be marked in the temporal structure of the utterance by the speeding up or slowing down of syllables. In the speech flow of people without disabilities, these structuring elements tend to occur at specific places (Zellner 1992, for review of literature see Zellner 1994 and Zellner 1998).

Psycholinguistic tests for English and French (Gee and Grosjean 1983; Keller et. al 1993; Zellner 1994, 1996, 1997a, 1997b, 2002) have shown that phrasing can only partially be predicted from syntax. A more adequate prediction is achieved with a psycholinguistically motivated model that splits up sentences into rhythmically balanced phrases. This means that phrases tend to be of similar length and can hardly ever be longer than a certain number of syllables. This is an assumption that seems quite plausible, considering cognitive and articulatory constraints. For French, this number rarely exceeds 12 (Zellner 1997b, 1998), and according to our data the same holds for German. That does not mean that syntactic and psycholinguistic aspects are in complete disaccord, but where the principles are in conflict, the psycholinguistic model is often more adequate (Zellner 1997a).

Based on statistic tests of these psycholinguistic principles Keller and Zellner Keller developed the following word grouping algorithm for French (Keller et al. 1993; Keller and Zellner 1996), which was refined for different speech rates (Zellner 1998).

A minor phrase boundary is inserted after every lexical word (noun, verb, adjective, adverb) that is followed by a grammatical word (preposition, conjunction, determiner…). For a certain number of special phenomena (fixed expressions, negations, complex verbal expressions, etc.), there are a few additional rules, which can insert additional boundaries or move existing ones.

Next major phrase boundaries are set, first of all at punctuation marks. In longer sentences, where the resulting phrases between major boundaries exceed the critical number of syllables, additional major boundaries are introduced. In fact, the minor boundary that is closest to the middle of the sentence is upgraded to major boundary status, and this procedure is iterated for each phrase between major boundaries until no such phrase is longer than twelve syllables.

3.2. Temporal Calculation

For the calculation of the segment durations, we rely mainly on statistic models built on manually labelled corpora. They are general linear models (GLM) that use input parameters such as the durational class of the current segment and the surrounding segments, the syllabic structure, the grammatical status of the word, the position of the segment in the word and in the prosodic phrase (Keller et al. 1993; Keller and Zellner 1996; Zellner 1998).

In addition, a number of qualitative and serial position components have been identified for French and English (Zellner 1998; Keller et al. 2000; Zellner Keller 2001, 2002; Zellner Keller and Keller 2001).

3.3. Intonation

Intonation is calculated with a superpositional Fujisaki model (first presented in Fujisaki and Hirose 1982). The model implements relatively slow phrase commands to determine the general intonation contour in a prosodic phrase on the one hand, and relatively fast accent commands on the other. The resulting curves of both kinds of commands are summed up and result in the final f0 contour. The position, duration and slope of these commands have been obtained from a copy-synthesis-analysis and statistical classification of the respective parameters.

3.4. Adaptation to German

Although globally speaking, the LAIP model for French could be adapted without major difficulty to standard German, relatively minor changes had to be made at every level. Most of them seem to be language-specific; they can therefore provide some hints about multilingual aspects of prosody.

In the phrasing model, changes with respect to the model for French mainly concerned the definition of special situations where minor phrase boundaries are introduced or moved. As they mostly concern fixed expressions, it is not surprising that these definitions are language-specific. A major change in the algorithm for German is the introduction of an additional minor phrase boundary in front of the inflected verb form which, compared to other minor boundaries, tends to be upgraded to major phrase boundary status in longer sentences.

The temporal calculation model for German needed some changes, which are mainly due to phonological differences with respect to French. First of all, German has – due to the phonological differenciation of long and short vowels and the occurrence of syllabic consonants – a larger segment inventory than French. Moreover, German has lexical stress, which is not assumed for French. The differences of the two models are described in detail in Siebenhaar et al. (2001).

The intonation model of the German synthesis system keeps the general superpositional approach used for French, but as the notion of stress differs considerably between the two languages, some changes had to be made. The main difference lies in the consideration of lexical stress for the German system according to the traditional Fujisaki approach (Fujisaki and Hirose 1982), whereas the French system considers every syllable for accent commands. For the alignment of segmental and intonational parameters we use an adaptation of the approach defined by Mixdorff (1998).

4. Speech Synthesis for Dialectal Variants

This model will now be adapted to a synthesis of Swiss German dialects. As we are just at the beginning of our project, we have hardly any results that go beyond some statistical analyses as a base to elaborate the models. Therefore we will just mention some of the major problems that are to be resolved before initiating the elaboration of the models itself.

As current speech synthesis systems can hardly handle more than one style of speech, the first question is: Which style of speech is to be modelled? The models for French and standard German are based upon an analysis of read speech. In fact, they reproduce mainly a news reading style, which is considered as neutral. This approach is quite evident for written standard languages. The Swiss German dialects, however, are not standardized, which means that they are hardly written and read, and that they are, while the main variety for numerous programmes on radio or TV, rarely the language of the news.[4] Thus, not only will we build synthesis systems for new varieties, but we will also have to model a different style of speech, which probably has a considerably greater inherent variation in prosodic terms than the 'neutral' news reading style of standard German.

The style we have decided to analyze in order to build our synthesis model for Swiss German dialects is the one of public interviews. Its main advantage is to be naturally embedded in a communication situation while, at the same time, being characterized by a certain degree of formality that prevents too exotic prosodic patterns. Moreover, it is very common to conduct interviews in the local dialect in German speaking Switzerland, where in contrast to most other German speaking areas, the dialect has a high prestige and is even used by officials in formal public situations. In conclusion, the interview style seems to be the one that combines best naturalness and a certain degree of formal control of the language.

Nevertheless, even in the relatively formal interview situation, typical phenomena of spontaneous speech are observed. Apart from silent pauses, there are non-pathologic influences such as hesitations, filled pauses (*mm*, *aa* and prolongation of sounds, mostly vowels but also nasals, liquids and fricatives, e.g. *dassss*) and repetitions (*wo wo*). Apart from that, there are also quite a number of ungrammatical utterances, ellipses and unfinished sentences, as well as forms of self-repair, where speakers stop mid-turn and change some aspect of what they have just been saying.

All these observations suggest that the model to be built might differ to a greater degree from its predecessors. However, work on the temporal structuring at different speech rates (Zellner 1998) as well as first preliminary tests in different speaking styles have shown that phrase boundaries are the same across varieties, even if the means used to mark them may differ. These findings let us expect that the general approach applied in the LAIPTTS systems can justifiably be adopted to model these new data, too. Moreover, the LAIPTTS word grouping algorithm is robust enough to work reasonably well even with ungrammatical input.

One interesting aspect of choosing the same approach for the development of speech synthesis systems of different languages, dialects and styles is the fact that we can more or less directly compare the resulting models with respect to the parameters turning out to be significant and to their respective importance. The observations concerning the models themselves can then give hints as to the relevance of the different levels of prosody – hints that may be confirmed or not by perceptual tests.

The particularity of this project is that Bernese and Zurich German are so closely related that it is reasonable to elaborate perfectly parallel prosodic models. If in spite of that there are salient differences in the relevance of the chosen parameters, this may give us quite detailed information about the way in which prosody can vary independently of the segmental information and thus about the extent to which prosody is language- or dialect-specific respectively.

5. Outlook to a Dialectal Synthesis

In a speech synthesis system, each prosodic parameter can be modified independently of all others, all while maintaining the interactions between different aspects of prosody. This makes speech synthesis an invaluable tool for testing hypotheses concerning the relevance of these different aspects for the naturalness of speech as well as claims as to their language-specific or language-independent nature. Sound examples can be generated with different models at each linguistic level, and by submitting them to perception tests we can evaluate the importance of the different models one by one.

For instance, a text with lexical, syntactical and morphological information characteristic of the Zurich dialect can be synthesized with the Bernese prosodic model, or with a Bernese timing model, but a Zurich model for intonation. The misfittings shown in the examples with manually changed prosody will not arise. The resulting signal is then presented to native speakers of Swiss German, who are asked to give their opinion concerning the origin of the virtual speaker, the acceptability of the speech signal, etc. By doing this for all possible combinations of Zurich and Bernese models one can test to what extent word grouping, segment durations and f0 contours obey the same or different regularities in the two dialects, if intermediate or mixed models result in an acceptable output, etc.

Given the complex interrelations between prosodic aspects, the great advantage of speech synthesis compared to conventional analytic methods is that it will make a difference between prosodic phenomena that mainly depend on different prosodic structures of the languages/dialects, and such prosodic phenomena that do indeed depend on structures of another type, for example different segmental structure, different number of syllables… It will be very interesting, for example, to test whether different f0 contours of the Zurich sentence *S gaat guet* 'it goes well' and its Bernese equivalent *S geit guet* can be explained by the phonetic segmental input or whether they really reflect a different behaviour of the two dialects in the prosodic domain.

Since speech synthesis allows us to manipulate the different parameters consistently, changing the phrasing information does have an influence on timing aspects and timing aspects are in turn reflected in the f0 contour. The following example may give an impression of what we plan for a dialectal speech synthesis. We have produced a sound file from a French transcription with French phonetics, but with German prosodic rules. The output was generated with a French synthetic voice. So in effect, we have a hypothetical German speaking French without accent at the phonetic level but with little knowledge of French prosodics. *Voici un texte français qui est produit avec les règles pour le rythme et l'intonation de l'allemand mais avec la voix française.* 'Here is a French text that is produced with the rules for rhythm and intonation of German, but with a French voice.' (Ex_11_frz_LAIPTTS_D.wav). The f0 curve and the rhythmic structure on word level is presented in Figure 9. This sound file shows clear differences to the same sentence produced with the French prosodic rules (Ex_12_frz_LAIPTTS_F.wav) given in Figure 10. The differences becomes very clear when listening just to the first part of the sentence (Ex_13_Dvoici.wav and Ex_14_Fvoici.wav).

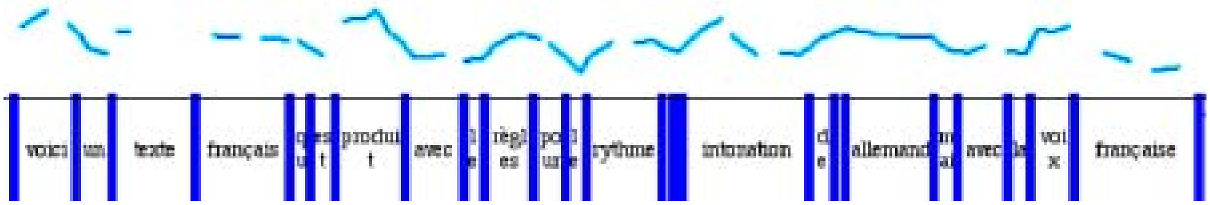

Figure 9: French text with German prosody (Ex_11_frz_LAIPTTS_D.wav)

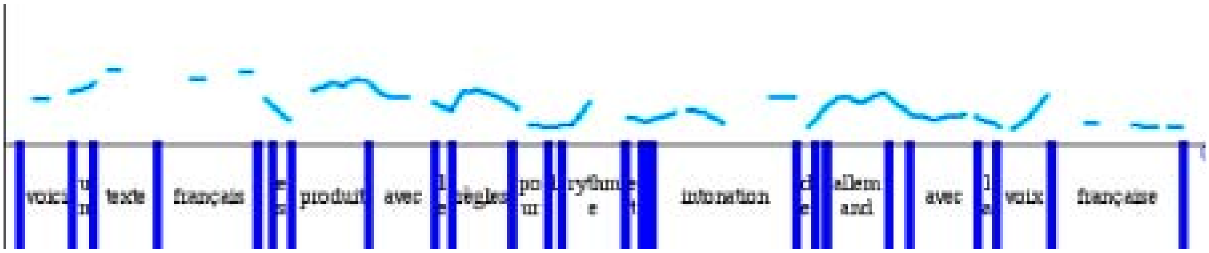

Figure 10: French text with French prosody (Ex_12_frz_LAIPTTS_F.wav)

These examples demonstrate the possibilities of speech synthesis systems to produce spoken text with different prosodic models respecting the different prosodic parameters. Thus, speech synthesis can be a way out of the dilemma of manipulating interrelated aspects of prosody. Let us see where this will take us.

With this synthesis tool we hope to obtain answers about the general structure of the prosody of the two dialects of Zurich and Berne, not about specific patterns in specific communicative functions. And while building the system we will no doubt detect the black holes in our knowledge of prosody, which we will have to fill with information obtained through new analyses. At the end we will be able to generate sound examples comparable to those presented above with different prosodic models that can be submitted to perception tests. These tests may provide valuable hints about the importance of different phonetic and prosodic aspects in dialect classification by naive listeners.

Acknowledgements: Many thanks to Brigitte Zellner Keller for her many ideas and critical review. This research is supported by the Swiss National Science Foundation.

References

Caelen-Haumont, G. 1991. *Stratégies des locuteurs et consignes de lecture d’un texte: Analyse des interactions entre modèles syntaxiques, sémantiques, pragmatique et paramètres prosodique.* Thèse de doctorat d’etat: Université d'Aix-en-Provence (unpublished).

Dieth, E. 1986. *Schwyzertütschi Dialäktschrift. Dieth-Schreibung*. 2nd edition, bearbeitet und herausgegeben von Christian Schmid-Cadalbert (Lebendige Mundart, 1). Aarau: Sauerländer.

Phonogrammarchiv der Universität Zurich (ed.) 2000. *Der sprechende Atlas."Gespräch am Neujahrstag" in 24 Dialekten (CD)*. Zurich: Phonogrammarchiv der Universität Zurich.

Fujisaki, H. and Hirose, K. 1982. 'Modelling the dynamic characteristics of voice fundamental frequency with applications to analysis and synthesis of intonation'. In *Preprints of the Working Group on Intonation, 13th International Congress of Linguists*. Tokyo. 57–70.

Gee, J. and Grosjean, F. 1983. 'Performance Structures: A Psycholinguistic and Linguistic Appraisal'. *Cognitive Psychology 15*: 411–458.

Gilles, P. and Schrambke, R. 2000. 'Divergenz in den Intonationssytemen rechts und links des Rheins. Die Sprachgrenze zwischen Breisach (Baden) und Neuf-Brisach (Elsass)'. In Funk, E., Konig, W. and Renn, M. (eds.), *Bausteine zur Sprachgeschichte. Referate der 13. Arbeitstagung zur alemannischen Dialektologie in Augsburg (29.9.–3.10.1999)*. Heidelberg: Winter. 87–98.

Haas, W. 2000. 'Die deutschsprachige Schweiz'. In Bickel, H. and Schläpfer, R. (eds.), *Die viersprachige Schweiz*. 2nd ed. (Reihe Sprachlandschaft, 25). Aarau, Frankfurt, Salzburg: Sauerländer. 57–138.

Hengartner, T. 1995. 'Dialekteinschätzung zwischen Kantonsstereotyp und Hörbeurteilung. Faktoren der Einschätzung schweizerdeutscher Dialekte'. In Loffler, H. (ed.), *Alemannische Dialektforschung. Bilanz und Perspektiven. Beitrage zur 11. Arbeitstagung alemannischer Dialektologen* (Basler Studien zur deutschen Sprache und Literatur 68). Tubingen/Basel: Francke. 81–95.

Hofer, L. 1997. *Sprachwandel im städtischen Dialektrepertoire. Eine variationslinguistische Untersuchung am Beispiel des Baseldeutschen* (Basler Studien zur deutschen Sprache und Literatur 72). Tubingen/Basel: Francke.

Keller, E. 1994. 'Fundamentals of phonetic science'. In E. Keller (ed.), *Fundamentals of Speech Synthesis and Speech Recognition: Basic Concepts, State of the Art, and Future Challenges*. Chichester: John Wiley. 5–21.

Keller, E. (to appear). La vérification d'hypothèses linguistiques au moyen de la synthèse de la parole. *Cahiers de l’institut de linguistique (Université de Louvain) 28.*

Keller, E., Zellner, B., Werner, S. and Blanchoud, N. 1993. 'The prediction of prosodic timing: Rules for final syllable lengthening in French'. In House, D. and Touati P. (eds.), *Proceedings ESCA Workshop on Prosody, September 27–29. Lund, Sweden* (Working Papers (Department of Linguistics and Phonetics, Lund University) 41). 212–215.

Keller, E. and Zellner, B. 1996. 'A timing model for fast French'. *York Papers in Linguistics 17*: 53–75.

Keller, E., & Zellner, B. 1998. 'Motivations for the prosodic predictive chain'. In *Proceedings of ESCA Symposium on Speech Synthesis*. Paper 76, pp. 137-141. Jenolan Caves, Australia. Available at www.unil.ch/imm/docs/LAIP/Kellerdoc.html.

Keller, E. and Zellner Keller, B. 2000. 'New Uses for Speech Synthesis'. *The Phonetician 81*: 35–40.

Keller, E., Zellner Keller, B. and Local, J. 2000. 'A serial prediction component for speech timing.' In Sendlmeier, W. F. (ed.), *Speech and Signals. Aspects of Speech Synthesis and Automatic Speech Recognition*. (Forum Phoneticum 69). Frankfurt am Main: Hector. 41–49.

Mixdorff, H. 1998. *Intonation Patterns of German – Model-based Quantitative Analysis and Synthesis of F0 contours*. http://www.tfh-berlin.de/usr1/doz/mixdorff/public_html/thesis.htm

Siebenhaar, B. and Wyler, A. 1997. *Dialect and high German in German-Speaking Switzerland*. Zurich: Pro Helvetia.

Siebenhaar B., Zellner Keller B. and Keller E. 2001. 'Phonetic and Timing Considerations in a Swiss High German TTS System'. In Keller, E., Bailly, G., Monaghan, A., Terken, J. and Huckvale, M. (eds.), *Improvements in Speech Synthesis*. Chichester: John Wiley. 165–175.

Werlen, I. 1980. 'R im Schweizerdeutschen'. *Zeitschrift für Dialektologie und Linguistik 47*: 52–76.

Werlen, I. 1985. 'Zur Einschätzung von schweizerdeutschen Dialekten'. In Werlen, I. (ed.), *Probleme der schweizerischen Dialektologie. 2. Kolloquium der Schweiz. Geisteswissenschaftlichen Gesellschaft 1978*. Freiburg/Fribourg: Editions Universitaires. 195–266.

Zellner, B. 1992. 'Le bé bégayage et euh... l'hésitation en français spontané'. In *Actes des 19ème Journées d'études sur la Parole (J.E.P).* Bruxelles. 481–487.

Zellner, B. 1994. 'Pauses and the temporal structure of speech'. In Keller, E. (ed.), *Fundamentals of speech synthesis and speech recognition*. Chichester: John Wiley. 41–62.

Zellner, B. 1996. 'Structures temporelles et structures prosodiques en français lu'. *Revue Française de Linguistique Appliquée 1*: 7–23.

Zellner, B. 1997a. 'Improving Speech Fluency in French through Psycholinguistic Principles'. In Borchardt, F. L. and Johnson, E.M.T. (eds.), *14th CALICO Annual Symposium*, [CD-ROM]. Durham: CALICO.

Zellner, B. 1997b. 'Fluidité en synthèse de la parole'. In Keller, E. and Zellner, B. (eds.), *Les défis actuels en synthèse de la parole* (Etudes des Lettres 3). Lausanne: Université de Lausanne. 47– 78.

Zellner, B. 1998. *Caractérisation et prédiction du débit de parole en français. Une étude de cas.* Thèse de Doctorat. Faculté des Lettres, Université de Lausanne. http://www.unil.ch/imm/docs/LAIP/pdf.files/Zellner_Dissertation.pdf

Zellner Keller, B. 2001. 'La modélisation du rythme de parole. Le problème de la cohérence temporelle'. In *Oralité et Gestualité. Interactions et comportements multimodaux dans la communication*. Actes. Aix en Provence (France). 640 –645.

Zellner Keller B. 2002. 'Revisiting the Status of Speech Rhythm'. In Bel B. and Marlien I. (eds.), *Proceedings of the Speech Prosody 2002 conference, 11–13 April 2002*. Aix-en-Provence: Laboratoire Parole et Langage. 727–730.

Zellner Keller B. and Keller E. (2001). 'A non linear rhythmic component in various styles of speech'. In Keller, E., Bailly, G., Monaghan, A., Terken, J. and Huckvale, M. (eds.), *Improvements in Speech Synthesis*. Chichester: John Wiley. 284–291

[1] Praat 4.0: http://www.fon.hum.uva.nl/praat/

[2] Some of these problems can be named here: Speech synthesis systems still have a sound quality that is far from being mistakable for a human, they obey to rigid rules, they are inflexible, they are not interactive, there is no contextualisation, they miss emotionality (cf. Keller to appear).

[3] There is a French and German online synthesis at http://www.unil.ch/imm/docs/laip. For offline use both systems can be downloaded for free from the same website.

[4] Local radio stations, where the dialect is the main variety, have recently developed a specific dialectal news reading style. In contrast to the one practised in the standard language, however, it aims to appear as spontaneous as possible. The speaker is supposed to hide the fact that s/he is actually reading, as this might be perceived as unnatural in the dialect. For a general view of the linguistic situation of German speaking Switzerland and its diglossia see Siebenhaar, Wyler (1997), for the historical background see Haas (2000).