HMM-based Speech Synthesis with Various Degrees of Articulation: a Perceptual Study

Benjamin Picart, Thomas Drugman, Thierry Dutoit

TCTS Lab, Faculté Polytechnique (FPMs), University of Mons (UMons), Belgium

Abstract. HMM-based speech synthesis is very convenient for creating a synthesizer whose speaker characteristics and speaking styles can be easily modified. This can be obtained by adapting a source speaker's model to a target speaker's model, using intra-speaker voice adaptation techniques. In this article, we focus on high-quality HMM-based speech synthesis integrating various degrees of articulation, and more specifically on the internal mechanisms leading to the perception of the degrees of articulation by listeners. Therefore the process of adapting a neutral speech synthesizer to generate hypo and hyperarticulated speech is broken down into four factors: cepstrum, prosody, phonetic transcription adaptation as well as the complete adaptation. The impact of these factors on the perceived degree of articulation is studied. Moreover, this study is complemented with an Absolute Category Rating (ACR) evaluation, allowing the subjective assessment of hypo/hyperarticulated speech through various dimensions: comprehension, non-monotony, fluidity and pronunciation. This article quantifies the importance of prosody and cepstrum adaptation as well as the use of a Natural Language Processor able to generate realistic hypo and hyperarticulated phonetic transcriptions.

1 Introduction

The "H and H" theory [1] proposes two degrees of articulation of speech: hyperarticulated speech, for which speech clarity tends to be maximized, and hypoarticulated speech, where the speech signal is produced with minimal efforts. Therefore the degree of articulation provides information on the motivation/personality of the speaker vs the listeners [2]. Speakers can adopt a speaking style that allows them to be understood more easily in difficult communication situations. The degree of articulation is characterized by modifications of the phonetic context, of the speech rate and of the spectral dynamics (vocal tract rate of change). The common measure of the degree of articulation consists in defining formant targets for each phone, taking coarticulation into account, and studying differences between real observations and targets vs the speech rate. Since defining formant targets is not an easy task, Beller proposed in [2] a statistical measure of the degree of articulation by studying the joint evolution of the vocalic triangle area (i.e. shape formed by vowels in the $F1$ - $F2$ space) and the speech rate.

We focus on the synthesis of different speaking styles, with a varying degree of articulation: neutral, hypoarticulated (or casual) and hyperarticulated (or clear) speech. “Hyperarticulated speech” refers to the situation of a speaker talking in front of a large audience (important articulation efforts have to be made to be understood by everybody). “Hypoarticulated speech” refers to the situation of a person talking in a narrow environment or very close to someone (few articulation efforts have to be made to be understood). It is worth noting that these three modes of expressivity are neutral on the emotional point of view, but can vary amongst speakers, as reported in [2]. The influence of emotion on the articulation degree has been studied in [3] and is out of the scope of this work.

Hypo/hyperarticulated speech synthesis has many applications: expressive voice conversion (e.g. for embedded systems and video games), “reading speed” control for visually impaired people (i.e. fast speech synthesizers, more easily produced using hypoarticulation), learning new languages: starting from hyperarticulated speech (high intelligibility, low speech rate, ...), the difficulty of the learning process could be increased when moving to hypoarticulated speech (low intelligibility, fast speech rate, ...), ...

This paper is in line with our previous works on expressive speech synthesis. The analysis and synthesis of hypo and hyperarticulated speech, in the framework of Hidden Markov Models (HMMs), has been performed in [4]. Significant differences between the three degrees of articulation were shown, both on acoustic and phonetic aspects. We then studied the efficiency of speaking style adaptation as a function of the size of the adaptation database [5]. Speaker adaptation [6] is a technique to transform a source speaker’s voice into a target speaker’s voice, by adapting the source HMM-based model (which is trained using the source speech data) with a limited amount of target speech data. The same idea lies for speaking style adaptation [7] [8]. We were therefore able to produce neutral/hypo/hyperarticulated speech directly from the neutral synthesizer. We finally implemented a continuous control (tuner) of the degree of articulation on the neutral synthesizer [5]. This tuner was manually adjustable by the user to obtain not only neutral/hypo/hyperarticulated speech, but also any intermediate, interpolated or extrapolated articulation degrees, in a continuous way. Starting from an existing standard neutral voice with no hypo/hyperarticulated recordings available, the ultimate goal of our research is to allow for a continuous control of its articulation degree.

This paper focuses on a deeper understanding of the phenomena responsible in the perception of the degree of articulation. This perceptual study is necessary as a preliminary step towards performing a speaker-independent control of the degree of articulation. Indeed the articulation degree induces modifications in the cepstrum, pitch, phone duration and phonetic transcription. In this work, these modifications are analyzed and quantified in comparison with a baseline, in which a straightforward, phone-independent constant ratio is applied to the pitch and phone durations of the neutral synthesizer in order to get as close as possible to real hypo/hyperarticulated speech. This perceptual study is complemented with an evaluation assessing the hypo/hyperarticulated speech quality through various dimensions: comprehension, non-monotony, fluidity and pronunciation.

After a brief description of the contents of our database in Section 2, the implementation of our synthesizers in the HMM-based Speech Synthesis System HTS ("H-Triple-S" - a toolkit publicly available in [9]) is detailed in Section 3. Results with regard to effects influencing the perception of the degree of articulation are given in Section 4. Finally Section 5 concludes the paper.

2 Database

For the purpose of our research, a new French database was recorded in [4] by a professional male speaker, aged 25 and native French (Belgium) speaking. The database contains three separate sets, each set corresponding to one degree of articulation (neutral, hypo and hyperarticulated). For each set, the speaker was asked to pronounce the same 1359 phonetically balanced sentences (around 75, 50 and 100 minutes of neutral, hypo and hyperarticulated speech respectively), as neutrally as possible from the emotional point of view. A headset was provided to the speaker for both hypo and hyperarticulated recordings, in order to induce him to speak naturally while modifying his articulation degree [4].

3 HMM-based Speech Synthesis

3.1 Conception of the Speech Synthesizers

An HMM-based speech synthesizer [10] was built, relying on the implementation of the HTS toolkit (version 2.1). 1220 neutral sentences sampled at 16 kHz were used for the training, leaving around 10% of the database for synthesis. For the filter, we extracted the traditional Mel Generalized Cepstral (MGC) coefficients (with $\alpha = 0.42$, $\gamma = 0$ and order of MGC analysis = 24). For the excitation, we used the Deterministic plus Stochastic Model (DSM) of the residual signal proposed in [11], since it was shown to significantly improve the naturalness of the delivered speech. More precisely, both deterministic and stochastic components of DSM were estimated on the training dataset for each degree of articulation. In this study, we used 75-dimensional MGC parameters (including Δ and Δ^2). Moreover, each covariance matrix of the state output and state duration distributions were diagonal.

In order to quantify the effects influencing the perception of the degree of articulation, we applied intra-speaker voice adaptation techniques (in the same spirit of inter-speaker voice adaptation) on this neutral HMM-based model in order to obtain directly a hypo/hyperarticulated model. Speaker adaptation is a technique to obtain a target speaker's voice model from a source speaker's voice model, using a limited amount of target speaker's speech data. The source speaker's voice model should be trained using a large number of source speaker's speech data in order to obtain a reliable model, from which adaptation could be performed. Proceeding this way allowed us to decompose each step of the adaptation process, to quantify the impact of each step on the listener perception of the degree of articulation.

Therefore, as illustrated in Figure 1, for each degree of articulation, this neutral HMM-based speech synthesizer was adapted using the Constrained Maximum Likelihood Linear Regression (CMLLR) transform [12] [13] in the framework of the Hidden Semi Markov Model (HSMM) [14], with hypo/hyperarticulated speech data to produce a hypo/hyperarticulated speech synthesizer. The linearly transformed models are further updated using a Maximum A Posteriori (MAP) adaptation [6].

In traditional HMM-based speech synthesis, the probability density functions of state durations are modeled by the state self-transition probabilities. Moreover, state duration models are only used for generating speech parameter sequences. As pointed out in [15], the expectation step of the EM algorithm is thus inconsistent. In HSMM-based speech synthesis, state duration distributions are modeled explicitly, allowing in this way a better representation of the temporal structure of human speech. HSMM has also the advantage of incorporating state duration models explicitly in the expectation step of the EM algorithm. Finally, HSMM is more convenient during the adaptation process to simultaneously transform both state output and state duration distributions.

MLLR adaptation is the most popular linear regression adaptation technique. The mean vectors and covariance matrices of state output distributions of the target speakers model are obtained by linearly transforming the mean vectors and covariance matrices of state output distributions of the source speaker's model [16]. The same idea lies for CMLLR. While MLLR is a model adaptation technique, CMLLR is a feature adaptation technique. In a model adaptation technique, a set of linear transformations is estimated to shift the means and alter the covariances in the source speaker's model so that each state in the HMM system is more likely to generate the adaptation data. In a feature adaptation technique, a set of linear transformations is estimated to modify the feature vectors in the source speaker's model so that each state in the HMM system is more likely to generate the adaptation data.

In the following, the full data models refer to the models trained on the entire training sets (1220 sentences, respectively neutral, hypo and hyperarticulated), and the adapted models are the models adapted from the neutral full data model, using hypo/hyperarticulated speech data. We showed in [5] that good quality adapted models can be obtained when adapting the neutral full data model with around 100-200 hypo/hyperarticulated sentences. On the other hand, the more adaptation sentences, the better the quality independently of the degree of articulation. This is why we chose in this work to adapt the neutral full data model using the entire hypo/hyperarticulated training sets. This will also allow us to remove from our results the amount of adaptation data from the possible perceptual effects (as it is studied in [5]).

3.2 Effects Influencing the Degree of Articulation

Based on the full data models and on the adapted models, four synthesizers are created: one for each effect to be analyzed, as summarized in Table 1. Through

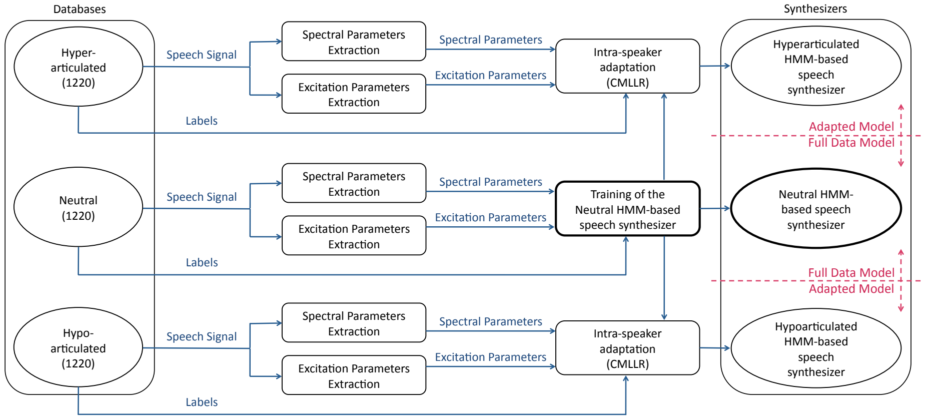
Databases
Hyper-articulated (1220)
Neutral (1220)
Hypo-articulated (1220)
Speech Signal
Labels
Spectral Parameters Extraction
Excitation Parameters Extraction
Spectral Parameters
Excitation Parameters
Intra-speaker adaptation (CMLLR)
Training of the Neutral HMM-based speech synthesizer
Synthesizers
Hyperarticulated HMM-based speech synthesizer
Neutral HMM-based speech synthesizer
Hypoarticulated HMM-based speech synthesizer
Adapted Model
Full Data Model

Fig. 1. Conception of the speech synthesizers.

the experimental evaluation described in Section 4, these four synthesizers will allow us to answer the following questions:

- Effect 1: Does simply applying a ratio on pitch and phone duration (while not adapting the cepstrum) sound like hypo/hyperarticulated speech? The first synthesizer (*Case 1*) is our baseline system and corresponds to the neutral full data model, where a straightforward phone-independent constant ratio is applied to decrease/increase pitch and phone durations to sound like hypo/hyperarticulated speech. This ratio is computed once for all over the hypo/hyperarticulated databases (see Section 2) by adapting the mean values of the pitch and phone duration from the neutral style. The phonetic transcription is manually adjusted to fit the real hypo/hyperarticulated transcription.
- Effect 2: What is the effect of cepstrum (neutral vs hypo/hyper) on the perception of the degree of articulation? The second synthesizer (*Case 2*) is constructed by only adapting pitch and phone duration distributions from the neutral full data model. The phonetic transcription is the same as from original hypo/hyperarticulated recordings.
- Effect 3: What is the effect of the phonetic transcription (neutral vs hypo/hyper) on the perception of the degree of articulation? The third synthesizer (*Case 3*) is constructed by adapting cepstrum, pitch and phone duration probability density functions from the neutral full data model. The phonetic transcription is not manually adjusted to fit real hypo/hyperarticulated transcription.
- Effect 4: Will the complete adaptation improve the perception of the degree of articulation compared to previous cases? The last synthesizer (*Case 4*) is built by adapting cepstrum, pitch and phone duration distributions of the neutral full data model. The phonetic transcription is the same as from original hypo/hyperarticulated recordings.

Table 1. *Conception of four different synthesizers, each of them focusing on an effect influencing the degree of articulation.*

	Full Data Model (Neutral)				Adapted Model (Hypo/Hyper)			
	Cepstrum	Pitch	Duration	Phon. Transcr.	Cepstrum	Pitch	Duration	Phon. Transcr.
Case 1	X	Ratio	Ratio					X
Case 2	X					X	X	X
Case 3				X	X	X	X	
Case 4					X	X	X	X

4 Experiments

In order to assess the performance of our synthesizers, two separate subjective experiments are conducted. Section 4.1 is dedicated to the evaluation of the influence of each factor explained in Section 3.2 on the perception of the degree of articulation. Section 4.2 complements the first evaluation by performing an Absolute Category Rating (ACR) test on other perceptual aspects of the synthetic speech.

4.1 Evaluation of the Perceived Degree of Articulation

For this evaluation, listeners were asked to listen to three sentences: the two reference sentences A (neutral) and B (hypo/hyper) synthesized by the full data models; the test sentence X synthesized by one of the four synthesizers described in Table 1 (randomly chosen), which could be either hypo or hyperarticulated depending on the articulation of B. Then participants were given a continuous scale, ranging from -0.25 to 1.25. A and B were placed at 0 and 1 respectively. Given this, they were asked to tell where X should be located on that scale. Evaluation was performed on the test set, composed of sentences which were neither part of the training set nor of the adaptation set.

The test consisted of 20 triplets. For each degree of articulation, 10 sentences were randomly chosen from the test set. During the test, listeners were allowed to listen to each triplet of sentences as many times as wanted, in the order they preferred. However they were not allowed to come back to previous sentences after validating their decision. 24 people, mainly naive listeners, participated to this evaluation. The mean Perceived Degree of Articulation (PDA) scores, together with their 95% confidence intervals are shown in Figure 2. The closer to 1 the PDA scores, the better the synthesizer as it leads to an efficient rendering of the intended degree of articulation.

From this figure, we clearly see the advantage of using an HMM to generate prosody (pitch and phone duration) instead of applying a straightforward phone-independent constant ratio to the neutral synthesizer prosody, in order to get

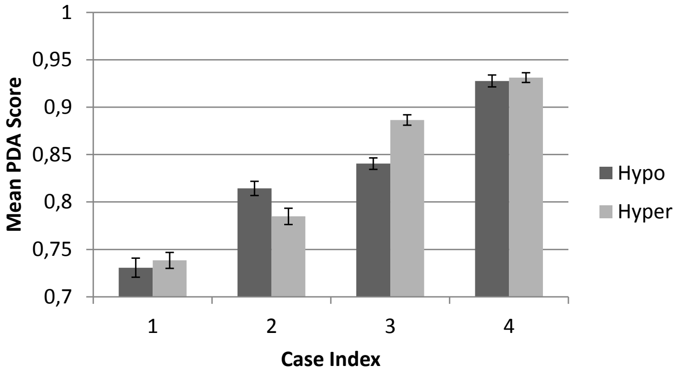

Fig. 2. Subjective evaluation - Mean PDA scores with their 95% confidence intervals (CI) for each degree of articulation.

as close as possible to real hypo/hyperarticulated speech (*Case 1* vs *Cases 2*, *3*, *4*).

The effects of cepstrum adaptation (*Case 2* vs *Case 4*) and phonetic adaptation (*Case 3* vs *Case 4*) are also highlighted. It can be noted that adapting the cepstrum has a higher impact on the rendering of the articulation degree than adapting the phonetic transcription (the gap between *Case 2* and *Case 4* is bigger than the gap between *Case 3* and *Case 4*). Moreover, it is also noted that this conclusion is particularly true for hyperarticulated speech, while the difference is less marked for hypoarticulation. Therefore *Case 2* indicates that the influence of spectral features is slightly more dominant for hyperarticulated speech. This might be explained by the fact that spectral changes (compared to the neutral style) induced by an hyperarticulation strategy are important to be modeled by the HMMs. Although significant spectral modifications are also present for hypoarticulated speech, it seems that their impact on the listener perception is marked to a lesser extent.

When analyzing *Case 3*, it is observed that a lack of appropriate phonetic transcription is more severe for hypoarticulated speech. Indeed, we have shown in [4] that hypoarticulated speech is characterized in particular by a high number of phone deletions, which is more important than the effect of phone insertions for hyperarticulated speech. This effect being stronger for hypoarticulated speech, we can easily understand that it will lead to a greater degradation of the speech signal perceived by the listeners.

Finally, it is noted that a high performance is achieved by the complete adaptation process (*Case 4* vs ideal value 1, which is the speech synthesized using the full data hypo/hyperarticulated models). This proves the efficiency of the degree of articulation CMLLR adaptation based on HMMs.

4.2 Absolute Category Rating Test

This experiment is based on the framework described in [17]. A Mean Opinion Score (MOS) test was complemented with an evaluation of various aspects of speech: comprehension, non-monotony, fluidity and pronunciation.

Table 2. *Question list asked to listeners during the ACR test, together with their corresponding extreme category responses.*

Test	Questions (Extreme Answers)
MOS	How did you appreciate globally what you just heard? (Very bad - Very good)
Comprehension	Did you find it difficult to understand the message? (Very difficult - Very easy)
Non-monotony	How would you characterize the speech intonation? (Very monotonous - Very varied)
Fluidity	How would you characterize the speech fluidity? (Very jerky - Very fluid)
Pronunciation	Did you hear some pronunciation problems? (Serious problems - No problem)

For this evaluation, 22 listeners were asked to listen to 20 test sentences, synthesized by one of the four synthesizers described in Table 1 (randomly chosen), which could be either hypo or hyperarticulated. These sentences were randomly chosen amongst the held-out set of the database (used neither for training nor for adaptation). Sentences were played one at a time. For each of them, listeners were asked to rate according to the 5 aspects cited above. Table 2 displays how the listeners were requested to respond. Listeners were given 5 continuous scales (one for each question to answer) ranging from 1 to 5 (these marks are associated with the extreme category answers in Table 2). These scales were extended one point further on both sides (ranging therefore from 0 to 6) in order to prevent border effects. During the test, listeners were allowed to listen to each sentence as many times as wanted. However they were not allowed to come back to previous sentences after validating their decision.

Mean scores are shown in figure 3. The MOS test shows an improvement in speech quality from *Case 1* to *Case 4*, for both hypo and hyperarticulated speech. This proves again the efficiency of the CMLLR adaptation process for producing high-quality synthetic speech. We clearly see an increase (decrease) in the intelligibility of hyper (hypo) articulated speech from *Case 1* to *Case 4* when analyzing the comprehension test. These results were expected considering our definition of hypo/hyperarticulated speech, and corroborate our findings of Section 4.1. The intelligibility of hyperarticulated speech is much higher for the complete adaptation process (*Case 4*) than for the baseline (*Case 1*).

A dramatic increase in monotony is observed for hypoarticulated speech (from *Case 1* to *Case 4*), while no significant variations were noticed for hy-

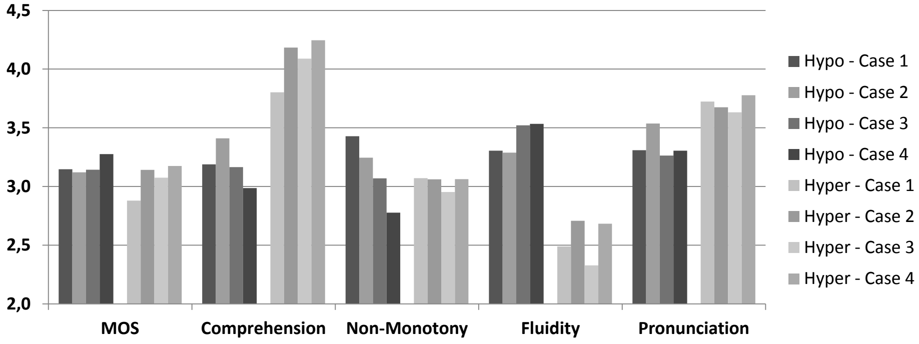

Fig. 3. Subjective evaluation - ACR test.

perarticulated speech. Going from *Case 1* to *Case 4* means getting closer to the target hypo or hyperarticulated speech. The sentence-wise intonation/variations, called the suprasegmental features, are reduced to the minimum in hypoarticulated speech because of the fastest speech rate, explaining the dramatic increase in monotony observed for hypoarticulated speech (from *Case 1* to *Case 4*). They could be enhanced in hyperarticulation because of the slowest speech rate. However, no significant differences are observed here, because the suprasegmental features were not amplified by our speaker from the neutral style to the hyperarticulated one.

The fluidity test shows that hypoarticulated speech is more fluid that hyperarticulated speech. This is due to the fact that hypoarticulated speech is characterized by a lower number of breaks and glottal stops, shorter phone durations and higher speech rate (as proven in [4]). All these effects lead to an impression of fluidity in speech, while the opposite tendency is observed in hyperarticulated speech. This explains also the fact that starting from our baseline (*Case 1*) and moving to the target hypo and hyperarticulated speaking styles, the speech becomes respectively more or less fluid (albeit no progressive degradation of fluidity across cases is reported for hyperarticulated speech).

Surprisingly enough, *Case 2* gives the higher result in the comprehension and pronunciation tests for hypoarticulated speech. This means that in order to decrease the comprehension of a message, it is required to adapt cepstrum from the neutral style, so as to model the weaker articulatory efforts in hypoarticulated speech. In this latter case, formant targets will be marked to a lesser extent. Finally, hyperarticulated speech exhibits no significant pronunciation differences amongst the different cases.

5 Conclusions

This article aimed at analyzing the adaptation process, and the resulting speech quality, of a neutral speech synthesizer to generate hypo and hyperarticulated speech. The goal was to have a better understanding of the factors leading to high-quality HMM-based speech synthesis with various degrees of articulation (neutral, hypo and hyperarticulated). This is why adaptation was subdivided into four effects: cepstrum, prosody, phonetic transcription adaptation as well as the complete adaptation.

First the perceptual impact of these factors was studied through a Perceived Degree of Articulation (PDA) test. It was observed that an efficient prosody adaptation cannot be achieved by a simple ratio operation. It was also shown that adapting prosody alone, without adapting cepstrum highly degrades the rendering of the degree of articulation. The impact of cepstrum adaptation turned out to be more important than the effect of phonetic transcription adaptation. Besides, the importance of having a Natural Language Processor able to create automatically realistic hypo/hyperarticulated transcriptions has been emphasized. This evaluation also highlighted the fact that high-quality hypo and hyperarticulated speech synthesis requires the use of an efficient statistical adaptation technique such as Constrained Maximum Likelihood Linear Regression (CMLLR).

Secondly, an Absolute Category Rating (ACR) test was conducted in complement to the PDA evaluation. For hyperarticulated speech, it was observed that the more complete the adaptation process (in the sense of the PDA scores), the higher the quality and comprehension of speech. Nonetheless, no significant differences in monotony and pronunciation were found. Regarding hypoarticulated speech, Mean Opinion Score (MOS) scores and results of comprehension, monotony and fluidity were interestingly in line with the conclusions of the PDA test.

All audio examples used in the experimental evaluations of this study are available online at http://tcts.fpms.ac.be/~picart/.

Acknowledgments

Benjamin Picart is supported by the "Fonds pour la formation à la Recherche dans l'Industrie et dans l'Agriculture" (FRIA). Thomas Drugman is supported by the Walloon Region, SPORTIC project #1017095.

References

1. B. Lindblom, Economy of Speech Gestures, vol. The Production of Speech, Spinger-Verlag, New-York, 1983.
2. G. Beller, *Analyse et Modèle Génératif de l'Expressivité - Application à la Parole et à l'Interprétation Musicale*, PhD Thesis (in French), Universit Paris VI - Pierre et Marie Curie, IRCAM, 2009.
3. G. Beller, N. Obin, X. Rodet, *Articulation Degree as a Prosodic Dimension of Expressive Speech*, Fourth International Conference on Speech Prosody, Campinas, Brazil, 2008.
4. B. Picart, T. Drugman, T. Dutoit, *Analysis and Synthesis of Hypo and Hyperarticulated Speech*, Proc. Speech Synthesis Workshop 7 (SSW7), Kyoto, Japan, 2010.

5. B. Picart, T. Drugman, T. Dutoit, *Continuous Control of the Degree of Articulation in HMM-based Speech Synthesis*, Proc. Interspeech, Firenze, Italy, 2011.
6. J. Yamagishi, T. Nose, H. Zen, Z. Ling, T. Toda, K. Tokuda, S. King, S. Renals, *Robust Speaker-Adaptive HMM-based Text-to-Speech Synthesis*, IEEE Audio, Speech, & Language Processing, vol. 17, no. 6, pp. 1208-1230, August 2009.
7. J. Yamagishi, T. Masuko, T. Kobayashi, *HMM-based expressive speech synthesis – Towards TTS with arbitrary speaking styles and emotions*, Proc. of Special Workshop in Maui (SWIM), 2004.
8. T. Nose, M. Tachibana, T. Kobayashi, *HMM-Based Style Control for Expressive Speech Synthesis with Arbitrary Speaker's Voice Using Model Adaptation*, IEICE Transactions on Information and Systems, vol. 92, no. 3, pp. 489-497, 2009.
9. HMM-based Speech Synthesis System (HTS) website : http://hts.sp.nitech.ac.jp/
10. H. Zen, K. Tokuda, A. W. Black, *Statistical parametric speech synthesis*, Speech Commun., vol. 51, no. 11, pp. 1039-1064, 2009.
11. T. Drugman, G. Wilfart, T. Dutoit, *A Deterministic plus Stochastic Model of the Residual Signal for Improved Parametric Speech Synthesis*, Proc. Interspeech, Brighton, U.K., 2009.
12. V. Digalakis, D. Rtischev, L. Neumeyer, *Speaker adaptation using constrained reestimation of Gaussian mixtures*, IEEE Trans. Speech Audio Process., vol. 3, no. 5, pp. 357-366, 1995.
13. M. Gales, *Maximum likelihood linear transformations for HMM-based speech recognition*, Comput. Speech Lang., vol. 12, no. 2, pp. 75-98, 1998.
14. J. Ferguson, *Variable Duration Models for Speech*, in Proc. Symp. on the Application of Hidden Markov Models to Text and Speech, pp. 143-179, 1980.
15. H. Zen, K. Tokuda, T. Masuko, T. Kobayashi, T. Kitamura, *Hidden Semi-Markov Model Based Speech Synthesis System*, IEICE Transactions on Information and Systems, vol. 90, no. 5, 2007.
16. J. Yamagishi, T. Kobayashi, *Average Voice-based Speech Synthesis using HSMM-based Speaker Adaptation and Adaptive Training*, IEICE Transactions on Information and Systems, vol. 90, no. 2, 2007.
17. P. Boula de Mareüil, C. d'Alessandro, A. Raake, G. Bailly, M.-N. Garcia, M. Morel, *A joint intelligibility evaluation of French text-to-speech synthesis systems: the EvaSy SUS/ACR campaign*, Proc. LREC, pp. 2034-2037, Gênes, 2006.