

Variations prosodiques en synthèse par sélection d'unités : l'exemple des phrases interrogatives

Laurence Martin¹ Sophie Roekhaut^{2,3} Richard Beaufort^{2,3}

(1) Faculté de philosophie, arts et lettres

(2) Centre de traitement automatique du langage (CENTAL)

(3) Institut Langage et Communication

Université catholique de Louvain, Louvain-la-Neuve, Belgique

laurence.j.martin@student.uclouvain.be,

{sophie.roekhaut,richard.beaufort}@uclouvain.be

RÉSUMÉ

Cet article propose une méthode automatique d'augmentation des variations prosodiques en synthèse par sélection d'unités. Plus particulièrement, nous nous sommes intéressés à la synthèse de phrases interrogatives au sein du système de synthèse *eLite*, qui procède par sélection d'unités non uniformes et qui ne possède pas les unités nécessaires à la production de questions dans sa base de données. L'objectif de ce travail a été de pouvoir produire des interrogatives via ce système de synthèse, sans pour autant enregistrer une nouvelle base de données pour la sélection des unités. Après avoir décrit les phénomènes syntaxiques et prosodiques en jeu dans l'énonciation de phrases interrogatives, nous présentons la méthode développée, qui allie pré-traitement des cibles à rechercher dans la base de données, et post-traitement du signal de parole lorsqu'il a été généré. Une évaluation perceptive des phrases synthétisées via notre application nous a permis de percevoir l'intérêt du post-traitement en synthèse et de pointer les précautions qu'un tel traitement implique.

ABSTRACT

Prosodic variations in unit-based speech synthesis: the example of interrogative sentences

This paper proposes an automatic method to increase the number of possible prosodic variations in non-uniform unit-based speech synthesis. More specifically, we are interested in the production of interrogative sentences through the *eLite* text-to-speech synthesis system, which relies on the selection of non-uniform units, but does not have interrogative units in its speech database. The purpose of this work was to make the system able to synthesize interrogative sentences without having to record a new, interrogative database. After a study of the syntactic and prosodic phenomena involved in the production of interrogative sentences, we present our two-step method: an adapted pre-processing of the unit selection itself, and a post-processing of the whole speech signal built by the system. A perceptual evaluation of sentences synthesized by our approach is then described, which points out both pros and cons of the method and highlights some issues in the very principles of the *eLite* system.

MOTS-CLÉS : synthèse NUU, phrases interrogatives, variations prosodiques.

KEYWORDS: NUU text-to-speech synthesis, interrogative sentences, prosodic variations.

1 Introduction

De nos jours, la synthèse par sélection d'unités non-uniformes (*Non-Uniform Units*, NUU) reste la plus commercialisée. Ce succès, elle le doit au naturel de la parole qu'elle produit, résultat de son principe fondateur : choisir dans une base de données les unités de parole les plus proches de la mélodie à produire afin de les modifier le moins possible par traitement du signal. À toute médaille, cependant, son revers : en synthèse NUU, les variations prosodiques de la parole de synthèse sont limitées aux variations présentes dans la base de parole utilisée.

Plusieurs travaux ont d'ailleurs proposé d'enrichir les bases de données utilisées. Soit en enregistrant une base de données par style (émotion, expression, etc.) à produire : dans ce cas, le système choisit d'abord la base correspondant le mieux au style désiré, avant de réaliser la sélection des unités de parole (Kawanami *et al.*, 2000; Iida *et al.*, 2003). Soit en rassemblant les styles au sein d'une seule et même base : dans ce cas, l'étiquetage des unités de parole est enrichi de caractéristiques faisant référence au style, et l'algorithme de sélection est modifié pour en tenir compte (Strom *et al.*, 2006; Syrdal et Kim, 2008). L'enrichissement des bases a cependant deux inconvénients : la taille des bases obtenues, et la nécessité de disposer du même locuteur lorsque de nouveaux enregistrements sont nécessaires.

Afin de devoir éviter d'enrichir les bases de données, Roekhaut *et al.* (2010) ont proposé de modifier le système de synthèse lui-même, en intervenant en amont et en aval de la sélection des unités. En amont, en modifiant les *valeurs* de l'étiquetage à rechercher dans la base de données. En aval, en post-traitant le signal obtenu pour accentuer les caractéristiques prosodiques désirées. Le résultat est une parole effectivement expressive, mais parfois dégradée par le post-traitement réalisé.

Nous nous inscrivons dans la continuité directe des travaux de Roekhaut *et al.* (2010). Notre objectif est d'apporter des réponses aux questions que leurs résultats avaient suscitées. Pour ce faire, nous sommes partis d'un cas précis : celui de la synthèse de phrases interrogatives à partir d'une base de données de parole exclusivement déclarative. L'étude est donc spécifique, mais a été menée avec la volonté de proposer des résultats applicables à d'autres types de variations prosodiques.

La suite de cet article s'articule comme suit. Après avoir présenté en section 2 le système de synthèse concerné, nous analysons en section 3 le comportement prosodique des questions. Sur cette base, nous décrivons en section 4 les traitements mis en place pour synthétiser des interrogatives à partir d'unités déclaratives. Nous évaluons ensuite la méthode en section 5, et présentons, en section 6, les réflexions que nos résultats suscitent concernant l'étiquetage de la base de données du système.

2 eLite-LiONS

eLite, prononcé [i l a j t], est un système complet de synthèse de la parole à partir du texte développé à Multitel ASBL¹ de 2001 à 2008 et maintenu au CENTAL depuis. Le système comprend un module de traitement automatique du langage naturel (Beaufort et Ruelle, 2006), qui construit une représentation phonético-prosodique du texte, un module de sélection NUU, LiONS (Colotte et Beaufort, 2005), qui exploite cette représentation pour choisir dans une base de données les unités de parole à concaténer, et un module de traitement du signal, qui concatène les unités sélectionnées en se limitant à un lissage de leurs frontières par Copy-OLA (Bozkurt *et al.*, 2004).

L'algorithme LiONS. L'unité de sélection utilisée par LiONS est le diphone². La séquence de phonèmes de la phrase à prononcer est donc convertie en une séquence de diphones. Chaque diphone se voit associer une liste de critères de sélection linguistiques, qui sont calculés au niveau

1. Centre de recherche belge situé à Mons, Hainaut, Belgique.

2. Le diphone est une unité acoustique qui s'étend de la partie stable d'un phonème à la partie stable du phonème suivant. Cette unité englobe donc la phase de coarticulation entre phonèmes, si difficile à modéliser.

de la syllabe à laquelle il appartient. Les diphtonges aux frontières de syllabes ou de groupes reçoivent des caractéristiques particulières. L'ensemble « diphtonges-critères » constitue une cible, pour laquelle on recherche des candidats dans la base de données. Lorsque des candidats ont été trouvés pour chaque cible, la meilleure séquence de candidats est sélectionnée en optimisant un double coût « cible-concaténation ».

Le coût de concaténation est une mesure de la distance acoustique entre les candidats de deux cibles différentes amenées à être concaténées dans le signal de parole.

Le coût cible est la distance d'un candidat par rapport à sa cible et dépend des critères de sélection utilisés. L'objectif de ces critères est avant tout de permettre au système de distinguer les unités toniques et proéminentes des autres, atones et non-proéminentes. Pendant longtemps, cette distinction a été exclusivement réalisée sur la base de valeurs acoustiques : F0, durée, spectre et énergie (Black et Campbell, 1995; Balestri *et al.*, 1999). LiONS appartient à une deuxième génération de systèmes NUU, qui ont remplacé les critères acoustiques par des critères linguistiques afin d'autoriser plus de variations dans la courbe prosodique. Initialement, LiONS utilisait 40 critères linguistiques pour décrire une cible. De nombreux tests ont ensuite permis de réduire cette liste à 4 critères, tous calculés au niveau de la syllabe à laquelle appartient le diphtonge :

1. la structure syllabique : V, CV, VC, CVC, etc. (où V=voyelle et C=consonne) ;
2. l'accent syllabique : primaire (AP), secondaire (AS) ou non accentué (NA) ;
3. la position de la syllabe dans le groupe rythmique (GR). Le GR est une notion propre à eLite. Il s'agit d'un groupe de souffle portant un léger accent sur sa première syllabe (BOG), un accent marqué sur sa dernière syllabe (EOG) et susceptible d'être suivi d'une pause. Le GR est constitué d'une ou de plusieurs unités grammaticales ;
4. la position de la syllabe par rapport à la pause courte (SH), moyenne (MD) ou longue (LG). Une syllabe devant la pause est toujours proéminente, mais son contour intonatif varie selon le type de pause : légèrement montant devant SH et MD, il devient descendant devant LG.

L'exemple suivant illustre, sur un énoncé simple, les critères linguistiques calculés par LiONS à partir de l'analyse linguistique produite par eLite :

Analyse linguistique (eLite)	Mots	Aujourd'hui			,	il	fait	froid	.
	Syllabes	o''	ʒ u ʁ	d ɥ i'	_	i l	f ɛ'	f ʁ w a'	_
	GR	GR1				GR2			
Critères linguistiques (LiONS)	(1)	V	CVC	CV		VC	CV	CV	
	(2)	AS	NA	AP		NA	AP	AP	
	(3)	BOG		EOG		BOG		EOG	
	(4)			SH				LG	

3 Comportement prosodique des interrogatives

Typologie. Il existe de nombreuses typologies syntaxiques de la question. Dans le cadre de cette étude, nous avons décidé de rassembler les interrogatives selon les 4 classes suivantes :

1. Les questions partielles : l'interrogation porte sur un élément particulier de la phrase, qui est représenté par un mot interrogatif (*Où allons-nous ?*) ;
2. Les questions totales : l'interrogation porte sur la totalité de la phrase, qui appelle une réponse de type oui/non (*Vous avez bien dormi ?*) ;
3. Les questions alternatives : un choix entre plusieurs possibilités équivalentes et acceptables est proposé à l'interlocuteur (*Tu veux du thé ou du café ?*) ;
4. Les demandes de continuation : l'interrogation ne porte pas sur un élément de l'énoncé, mais pousse l'interlocuteur à poursuivre son développement (*Et alors ?*, *Ah bon ?*, *C'est-à-dire ?*).

Nous avons également réalisé un recensement des marqueurs syntaxiques de la question, et de la façon dont ils peuvent se combiner avec les quatre classes ci-dessus (voir figure 1). Le marquage syntaxique peut consister en :

- A. un mot (déterminant, adverbe ou pronom) interrogatif ;
- B. un « tag » en fin d'énoncé, par exemple *n'est-ce pas*, marquant une demande de confirmation établie par le locuteur (Grundstorm, 1973) ;
- C. la locution *est-ce que* ;
- D. une inversion sujet-verbe ;
- E. aucun marqueur : l'énoncé est déclaratif, mais présente un point d'interrogation à l'écrit.

Corpus. Sur cette base, nous avons constitué un corpus d'interrogatives pouvant être, de manière univoque, transcrites et ponctuées à l'aide d'un point d'interrogation. Afin de faciliter l'identification des comportements prosodiques propres à chaque classe de notre typologie, nous avons arrêté notre choix sur 123 énoncés relativement stéréotypés, provenant d'un CD audio d'exercices destinés à des apprenants du français langue seconde (Berthet *et al.*, 2006). Chaque énoncé a été manuellement classé dans notre typologie, puis a été transcrit phonétiquement et aligné avec le signal de parole au moyen de *Praat*, un logiciel libre d'annotation et de manipulation de données orales (Boersma et Weenink, 2011), et son module *EasyAlign* (Goldman, 2011). Les énoncés étudiés ont enfin été soumis à une analyse prosodique afin de produire un prosogramme (Mertens, 2004), représentation graphique du contour prosodique d'un énoncé, basée sur les valeurs de hauteur de chaque noyau syllabique exprimées en demi-tons. Ceci nous a permis d'observer, de manière systématique, les valeurs de hauteur accordées aux syllabes finales, ainsi qu'aux syllabes des éventuels marqueurs syntaxiques.

Analyse. Les comportements prosodiques observés sur notre corpus, illustrés sur les exemples de la figure 1, confirment plusieurs théories linguistiques (Delattre, 1966; Léon et Léon, 2007). En finale, seules les interrogatives de forme déclarative sont obligées de monter du fait de l'absence de tout marqueur syntaxique (Ex. 2.E, 4.E). À l'inverse, une montée en finale en présence d'un marqueur syntaxique se perçoit comme redondante et n'est pas obligatoire (Ex. marqués syntaxiquement par A, C ou D). Enfin, une montée est fréquente sur la dernière syllabe d'un mot interrogatif (Ex. marqués syntaxiquement par A), quelle que soit sa position dans l'énoncé. La littérature (Vion *et al.*, 2002; Fónagy, 2003) signale également que dans le cas d'une alternative, le premier terme serait montant et le second, descendant (3.C, 3.D, 3.E). Sans questions alternatives dans notre corpus, nous n'avons pu vérifier cette hypothèse. Par contre, notre corpus nous a permis d'observer que dans le cas d'énoncés marqués par un tag en finale, une descente et une courte pause précèdent toujours le tag, tandis qu'une montée prend place sur la dernière syllabe du tag (Ex. 2B).

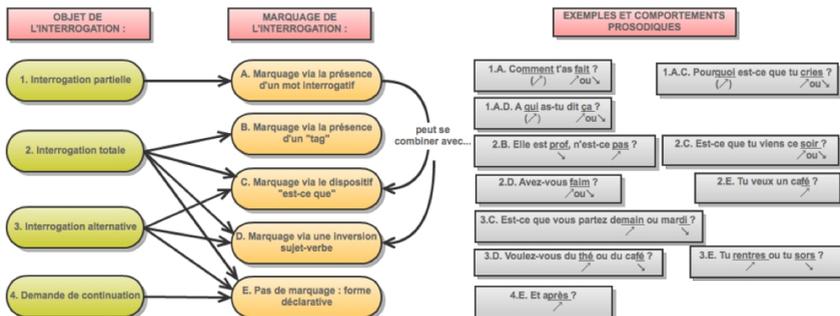


FIGURE 1 – Typologie syntaxique et prosodique des interrogatives

4 Traitements appliqués aux interrogatives

La base de données de parole manipulée par eLite-LiONS est constituée de 56 000 diphtonges provenant exclusivement de phrases déclaratives. Puisqu'elle ne contient pas d'unité proéminente et à contour mélodique ascendant en finale d'énoncé, cette base n'est donc pas adaptée telle quelle à la modélisation de certaines interrogatives, notamment celles qui ne sont pas marquées par la syntaxe et qui doivent alors obligatoirement monter en finale. Pour tendre vers l'interrogation, à l'instar de Roekhaut *et al.* (2010), nous intervenons par un pré-traitement en amont et un post-traitement en aval de la sélection des unités. Le pré-traitement doit permettre de choisir des unités dont la courbe prosodique, montante ou descendante, va dans le sens désiré. Le post-traitement doit accentuer cette tendance naturelle, la rendre audible et distinguable.

Pré-traitement. Le principe est de modifier les valeurs des critères linguistiques des cibles à rechercher dans la base de données. Selon le type de question, deux modifications peuvent s'envisager.

1) Soit, nous avons besoin d'une **intonation montante et proéminente** là où une déclarative serait naturellement plate ou descendante. Dans une phrase déclarative, les seules syllabes de ce type se situent à l'endroit d'une continuation majeure : une pause courte, correspondant dans le texte à une virgule. Nous avons forcé le système à sélectionner ces unités en imposant au critère linguistique « *distance par rapport à la pause* » la valeur « *devant pause courte* ». Cette modification a pu être appliquée à tous les cas où notre analyse prosodique a mis en évidence la nécessité d'une intonation montante, sauf à la dernière syllabe d'un mot interrogatif, parce que le mot interrogatif n'est jamais suivi d'une pause, contrairement à l'unité que nous aurions voulu lui substituer. Dans ce cas précis, seul le post-traitement décrit ci-dessous a pu être appliqué.

2) Soit, nous avons besoin d'une **intonation descendante** là où une déclarative serait naturellement ascendante. Dans une déclarative, cette intonation se retrouve typiquement en finale de phrase. Ici, nous avons forcé le système à sélectionner ces unités en imposant au critère linguistique « *distance par rapport à la pause* » la valeur « *devant pause longue* ». Ce cas ne concerne que les interrogatives terminées par un tag, dont la syllabe précédant le tag est caractérisée par une descente et une légère pause.

Post-traitement. L'analyse des prosogrammes de nos énoncés interrogatifs nous a permis de fixer à 3.4 demi-tons la différence de hauteur moyenne entre la syllabe montante de l'interrogative et la syllabe qui la précède. Pour obtenir cette différence de hauteur entre les syllabes concernées des interrogatives générées par eLite-LiONS, nous avons utilisé l'algorithme de synthèse PSOLA (Moulines et Charpentier, 1990) implémenté dans Praat.

La figure 2 illustre l'évolution de la courbe prosodique d'un énoncé subissant successivement les pré- et post-traitements décrits ci-dessus.

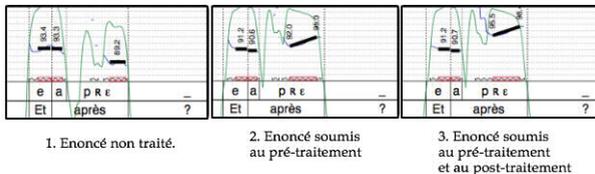


FIGURE 2 – Illustration de l'application des pré- et post-traitements

5 Évaluation

Vingt-et-un évaluateurs ont réalisé deux tests perceptifs en ligne³ afin d'évaluer trois aspects des phrases interrogatives produites : *le naturel du signal* (les traitements appliqués altèrent-ils la qualité du signal ?), *la force de l'intonation interrogative* (les questions générées sont-elles bien perçues comme telles ?) et, enfin, *le naturel de l'intonation interrogative* (le résultat prosodique obtenu est-il perçu comme naturel ?).

Naturel du signal :

Pour chacun des 6 énoncés proposés, les évaluateurs devaient indiquer s'ils préféraient avec ou sans post-traitement. Lorsque l'étape de pré-traitement a sélectionné des unités fortement proéminentes pour la dernière syllabe de l'énoncé, le post-traitement a accentué cette proéminence et provoqué une sortie du registre de la voix de la base de données : la majorité des participants a alors préféré les énoncés non post-traités. À l'inverse, lorsque l'étape de pré-traitement a été correctement réalisée, le post-traitement n'a pas été rejeté.

Force de l'intonation interrogative :

Les évaluateurs devaient détecter les 11 interrogatives de forme déclarative dans un ensemble de 16 énoncés comptant 5 vraies déclaratives. L'intonation La force d'interrogation de chaque énoncé a été évaluée sur une échelle de 1 à 5. Le test est concluant : la grande majorité des questions générées par notre application sont comprises comme telles. Seules 3 interrogatives sur 11 ont résisté à nos traitements : leur intonation normalement montante ne s'est pas montrée aussi marquée que prévu. Elles n'ont de ce fait pas été perçues comme interrogatives. Nous revenons sur les causes possibles de cette irrégularité en section 6.

Naturel de l'intonation :

Questions marquées syntaxiquement. Nous l'avons mentionné en section 3, une montée finale dans le cas de questions marquées syntaxiquement peut être perçue comme redondante. Dans ce cas précis, nous avons demandé aux évaluateurs de choisir entre 3 versions différentes des mêmes 6 questions marquées : une version sans traitement, une version avec pré-traitement uniquement, et une version avec pré- et post-traitement. Globalement, la majorité des évaluateurs préfèrent la version uniquement pré-traitée, dont la légère montée permet d'insister suffisamment sur la question, tout en évitant les éventuelles dégradations dues au post-traitement.

Questions marquées par un mot interrogatif. Nous rappelons que la montée normalement attendue sur les mots interrogatifs (section 3) ne peut être produite que par post-traitement (section 4). Au vu des résultats du test perceptif réalisé sur 22 énoncés, cette limite n'est cependant pas gênante : les 10 questions marquées par un mot interrogatif dont la mélodie n'était montante qu'en finale ont été jugées plus naturelles, ou sans différence perceptible avec les énoncés où la mélodie était également montante sur le mot interrogatif. Ce résultat s'explique sans doute par la difficulté que nous avons eue à produire une montée significative préalable sur les mots interrogatifs.

Questions taguées. Sur la base de notre corpus, nous avons constaté la nécessité d'une descente mélodique sur la dernière syllabe du mot précédant le tag, suivie d'une légère pause (section 3). Le test perceptif, réalisé sur 2 énoncés, a validé cette observation : des 2 courbes intonatives, les participants ont systématiquement préféré celle qui présentait un contour déclaratif descendant avant le tag. Le pré-traitement est donc pertinent.

3. Disponibles sur <http://cental.fltr.ucl.ac.be/testperceptif2/> et sur <http://cental.fltr.ucl.ac.be/testperceptif3/>.

Questions alternatives. Pour 3 énoncés, les évaluateurs ont départagé 2 intonations : la première, avec une montée sur le premier terme de la question uniquement, l'autre, avec une montée en finale également. La finale d'une question alternative ne correspond pourtant pas toujours au focus de la question : dans "*Il est en avril ou en septembre ton examen le plus difficile ?*", une montée finale semblerait mal venue. Cet exemple met au jour la nécessité de repérer le focus de l'interrogative *avant* d'en prédire la prosodie. Pourtant, la majorité des évaluateurs ont apprécié la montée en finale, même lorsqu'elle ne correspondait pas au focus. Ceci est peut-être dû au fait que le pré-traitement décrit est impossible à appliquer au premier terme des alternatives, qui ne profite de ce fait que du post-traitement, marquant alors moins bien la question.

Portée de l'ascendance finale. L'objectif était de déterminer le meilleur nombre de syllabes sur lequel réaliser la montée en finale. Les utilisateurs ont dû choisir entre deux versions de 6 énoncés : la première avec une montée sur la dernière syllabe, l'autre avec une montée sur les trois dernières syllabes. La majorité des réponses obtenues ne font aucune différence entre les deux intonations proposées, ou sont favorables à la montée sur la dernière syllabe uniquement. La montée sur les trois dernières syllabes n'a été préférée que dans le cas d'énoncés où le post-traitement a provoqué une sortie du registre de la voix de synthèse, du fait de la sélection en amont d'unités fortement proéminentes. La diffusion de la montée sur les trois dernières syllabes de l'énoncé permet alors, sans doute, de réduire cette proéminence exagérée.

6 Conclusions et perspectives

Les variations prosodiques de la synthèse par sélection d'unités sont par nature limitées aux variations présentes dans les bases de données de parole utilisées. C'est en partant de ce constat que de nombreux chercheurs ont proposé diverses méthodes pour enrichir les bases en question. Cependant, l'enrichissement des bases est coûteux, et lie dans le temps le système de synthèse à la disponibilité de la voix utilisée.

Afin d'éviter ces désagréments, et en partant du cas particulier des interrogatives, cet article a proposé une méthode alliant pré- et post-traitement de la sélection d'unités pour augmenter les possibilités prosodiques de la base. Le pré-traitement permet de choisir des unités présentant la tendance prosodique souhaitée, tandis que le post-traitement accentue cette tendance pour la rendre audible.

L'évaluation perceptive que nous avons réalisée a montré que dans l'ensemble, les variations prosodiques obtenues par notre approche étaient audibles, reconnaissables et naturelles. Cependant, l'évaluation a également mis au jour un point important : la qualité et la pertinence du post-traitement réalisé dépendent directement des caractéristiques acoustiques de l'unité traitée. Si l'unité à traiter est déjà aux limites du registre de la voix de la base de données, le post-traitement peut l'en faire sortir et dégrader le signal de manière audible. À l'inverse, si l'unité à traiter ne possède pas la tendance prosodique recherchée (ici, une montée ou une descente), le post-traitement n'a aucune efficacité.

Cette inefficacité du post-traitement a également été constatée lorsque l'unité à traiter, proéminente dans son contexte initial, ne l'est plus ou pas assez dans le contexte de la phrase de synthèse. Ce dernier constat est très important, parce qu'il remet en cause le bienfondé du recours à des critères purement linguistiques pour décrire les cibles à sélectionner. Ces critères sont-ils suffisants pour distinguer les unités qui, hors de leur contexte initial, conserveront un comportement prosodique donné ? Nous n'en sommes pas certains. Au contraire, ces résultats semblent indiquer que des valeurs acoustiques restent somme toute nécessaires pour obtenir un signal de parole où l'alternance entre syllabes proéminentes et non proéminentes respecte les standards de la langue.

Références

- BALESTRI, M., PACCHIOTTI, A., QUAZZA, S., SALZA, P. et SANDRI, S. (1999). Choose the best to modify the least : A new generation concatenative synthesis system. In *Proceedings of Eurospeech*, pages 2291–2294, Budapest, Hungary.
- BEAUFORT, R. et RUELLE, A. (2006). elite : système de synthèse de la parole à orientation linguistique. In *Proceedings of JEP*, pages 509–512, Dinard, France.
- BERTHET, A., HUGOT, C., KIZIRIAN, V., SAMPSONIS, B. et WAENDENDRIES, M. (2006). *Alter Ego 2*. Hachette.
- BLACK, A. et CAMPBELL, N. (1995). Optimising selection of units from speech databases for concatenative synthesis. In *Proceedings of Eurospeech*, pages 581–584, Madrid, Spain.
- BOERSMA, P. et WEENINK, D. (2011). Praat : doing phonetics by computer (version 5.2.28). <http://www.praat.org>.
- BOZKURT, B., DUTOIT, T., PRUDON, R., D’ALESSANDRO, C. et PAGEL, V. (2004). Chapter 1 : Reducing discontinuities at synthesis time for corpus-based speech synthesis. In NARAYANAN, S. et ALWAN, A., éditeurs : *Text To Speech Synthesis : New Paradigms and Advances*. Prentice Hall PTR.
- COLOTTE, V. et BEAUFORT, R. (2005). Linguistic features weighting for a text-to-speech system without prosody model. In *Proceedings of Interspeech*, pages 2549–2552, Lisbon, Portugal.
- DELATRE, P. (1966). Les dix intonations de base du français. *The French Review*, 40(1):1–14.
- FÓNAGY, I. (2003). Des fonctions de l’intonation : essai de synthèse. *Flambeau*, 29:1–20.
- GOLDMAN, J.-P. (2011). EasyAlign : an automatic phonetic alignment tool under Praat. In *Proceedings of Interspeech*, pages 3233–3236, Florence, Italy.
- GRUNDSTORM, A. (1973). L’intonation des questions en français standard. *Studia Phonetica* 8, pages 19–49.
- IIDA, A., CAMPBELL, N., HIGUCHI, F. et YASUMURA, M. (2003). A corpus-based speech synthesis system with emotion. *Speech Communication*, 40(1):161–187.
- KAWANAMI, H., MASUDA, T., TODA, T. et SHIKANO, K. (2000). Designing speech database with prosodic variety for expressive tts system. In *Proceedings of LRE*.
- LÉON, M. et LÉON, P. (2007). *La prononciation du français*. Armand Collin.
- MERTENS, P. (2004). Un outil pour la transcription de la prosodie dans les corpus oraux. *Traitement Automatique des langues*, 45(2):109–130.
- MOULINES, E. et CHARPENTIER, F. (1990). Pitch-synchronous waveform processing techniques for text-to-speech synthesis using diphones. *Speech communication*, 9(5):453–467.
- ROEKHAUT, S., GOLDMAN, J. et SIMON, A. (2010). A model for varying speaking style in tts systems. In *Proceedings of Speech Prosody*, Chicago, Illinois, USA.
- SCHRÖDER, M. (2001). Emotional speech synthesis : A review. In *Proceedings of Eurospeech*, pages 561–564, Aalborg, Denmark.
- STROM, V., CLARK, R. et KING, S. (2006). Expressive prosody for unit-selection speech synthesis. In *Proceedings of Interspeech*, Pittsburgh, Pennsylvania, USA.
- SYRDAL, A. et KIM, Y. (2008). Dialog speech acts and prosody : Considerations for tts. In *Proceedings of Speech Prosody*, pages 661–665, Campinas, Brazil.
- VION, M., COLAS, A. et al. (2002). La reconnaissance du pattern prosodique de la question : questions de méthode. *Travaux Interdisciplinaires Parole et Langage*, 21:153–177.