

Implantation d'un synthétiseur à formants de Klatt sur le processeur ST18940

Implementation of the Formant Klatt Synthesizer on the ST18940 DSP

M. OUADOU, A. RAJOUANI,
A. BELHADJ

LEESA, Faculté des Sciences, B.P. 1014,
Rabat, Maroc

P. BAYLOU et M. NAJIM

Equipe Signal et Image, ENSERB,
351, cours de la Libération,
F33405 Talence Cedex, France

RÉSUMÉ

Dans cet article, nous décrivons l'implantation du synthétiseur à formants de Klatt (série et parallèle) opérant en temps réel sur un seul processeur du signal ST8940 de SGS-Thomson. La simulation de l'algorithme de synthèse en virgule flottante a permis la détermination des conditions optimales de fonctionnement. La simulation en virgule fixe est réalisée pour l'estimation de la longueur adéquate du mot et la mise à l'échelle de l'ensemble des variables impliquées. La réalisation de ces deux étapes a

permis la programmation de l'algorithme de synthèse en langage assembleur du processeur. Notre objectif à court terme est la réalisation d'une carte compatible PC de synthèse de la parole autour du processeur ST18940.

Mots clés

Synthèse à formants, synthétiseur de Klatt, implantation temps réel, processeur de signal.

ABSTRACT

This paper describes the real time implementation of the Klatt speech formant synthesizer (cascade and parallel) on the SGS-Thomson DSP Processor. The determination of optimal functional conditions of the synthesis algorithm is performed by means of the floating point simulation. Fixed point simulation is realized to estimate the word length and to define the adequate scale of the used variables. The realization of these procedures allows the programming, in

assembly language, of the algorithm on the DSP. This is a first step towards a realization of a PC-compatible speech synthesis board based on the DSP processor.

KEY WORDS

Formant synthesis, Klatt synthesizer, real time implementation, digital signal processor.

1. Introduction

La technologie vocale a été considérablement modifiée par les progrès rapides de la technologie des semi-conducteurs et des techniques de codage. Toutefois le codage prédictif (LPC) et les formants restent les approches les plus utilisées pour la synthèse de la parole.

Les réalisations VLSI de synthétiseurs de parole (AMI536xx, SP-0256(GI), TMS51xx, TMS52xx, TMS50C4x, CNET-SDP186, ...) ont permis une production en série et à coût réduit de plusieurs produits [1, 2, 3, 4, 5].

Nous avions, en ce qui nous concerne, utilisé, en particulier, le TMS5220C pour la réalisation d'une carte périphé-

que de synthèse [3] et le TMS50C42 pour la réalisation, en collaboration avec le CNET, Texas-Instruments-France et la société ELAN-Informatique, d'une carte de synthèse multilingue compatible PC [4, 5].

La mise au point de circuits intégrant la synthèse par formants a été relativement lente. Les produits disponibles actuellement tels le MEA8000 et le PC2000 de Philips ou le CTI de Computer Consultants essayent de combler cette lacune alors que le synthétiseur de Klatt s'impose progressivement comme un standard dans les laboratoires de recherche. La situation s'est encore considérablement améliorée avec l'apparition des DSP : processeurs spécialisés en traitement du signal tels que les DSP16/32, TMS320, NEC77xx, AD21xx...

Des travaux antérieurs ont porté sur l'implantation de sous-ensembles, série ou parallèle du synthétiseur de Klatt sur le TMS32010 [18] ou le NEC7720 [18] ou le NEC7720 [19]. D'autres ont porté sur l'implantation du synthétiseur de Klatt série et parallèle sur plusieurs TMS32010 (DECtalk) ou NEC7720 (PROSE200). Les performances de PROSE2000 et de DECTalk (MITtalk [13]) pour l'Anglais ont donné leurs lettres de noblesse à la synthèse phonétique et au synthétiseur de Klatt série et parallèle. Ce synthétiseur a été utilisé par la suite pour la synthèse du Français [14, 15], de l'Allemand [16], de l'Espagnol [17] et de l'Arabe [8, 9].

Nous proposons l'implantation du synthétiseur de Klatt série et parallèle sur un seul processeur. Il est évident que cette structure nécessite un plus grand nombre de paramètres de contrôle et un temps de calcul plus grand.

Notre expérience sur le TS68930 (DSP de SGS-Thomson) [6, 7] nous a amenés à opter pour l'implantation du synthétiseur de Klatt série et parallèle sur un DSP plus récent et totalement compatible : le ST18940 [10]. Ce DSP intègre des concepts avancés dans le domaine des architectures tels que la structure Harvard, le « pipeline », le parallélisme et les possibilités de l'arithmétique double précision et complexe. La réalisation en temps réel du synthétiseur de Klatt entre dans le cadre de notre projet visant la mise au point d'un système intégré de synthèse par règles de langues sémitiques [8, 9].

La présentation du synthétiseur de Klatt et l'analyse des fonctions de simulation en virgule fixe est effectuée dans la section 2. La section 3 est consacrée à la discussion et à l'évaluation des résultats obtenus lors de l'implantation sur le DSP.

2. Simulation

La synthèse par formants est étroitement liée au processus physiologique de production de la parole [11, 12]. Lors de la propagation du son de la glotte aux lèvres, le spectre de l'excitation de la source est mis en forme par les résonances du conduit vocal appelées formants. L'objectif de la synthèse par formants est de produire un signal de parole à partir des informations sur les fréquences, les bandes passantes et dans certains cas sur les amplitudes relatives des formants. Les synthétiseurs à formants utilisent des

filtres tout-pôles du second ordre en configuration série, parallèle ou série et parallèle.

Le synthétiseur de Klatt série et parallèle [12] est représenté par la figure 1. Ce synthétiseur est composé de deux systèmes : l'ensemble des sources d'excitation et le système de filtrage. Le résonateur est un filtre du second ordre régi par une équation récurrente de la forme :

$$y(nT) = Ax(nT) + By(nT - T) + Cy(nT - 2T)$$

où $y(nT)$ est la sortie, $x(nT)$ est l'entrée et T la période d'échantillonnage. Les constantes A , B , C sont liées aux formants F et aux bandes passantes BP par les équations suivantes :

$$C = -\exp(-2\pi \cdot BP \cdot T),$$

$$B = 2 \cdot \exp(-\pi \cdot BP \cdot T) \cos(2\pi \cdot F \cdot T)$$

et

$$A = 1 - B - C.$$

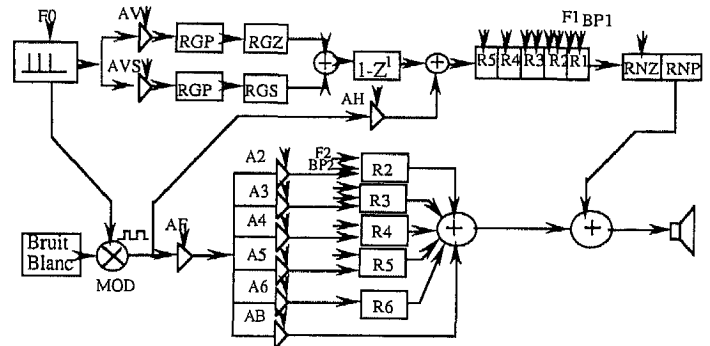


Fig. 1. — Structure du synthétiseur (R_i : résonateurs, F_i et B_i : formants et bandes passantes, A_i - AB : amplitudes, AV, AVS, AH, AF : gains des sources, F₀ : fréquence fondamentale).

La version que nous avons utilisée, nécessite 39 paramètres de contrôle dont 20 variables [12].

La simulation de l'algorithme de synthèse en virgule flottante permet l'étude détaillée de son comportement et par conséquent une première évaluation des conditions nécessaires à son fonctionnement en précision finie. Le programme de simulation est interactif et permet l'analyse de plusieurs fonctions, en particulier :

- la sensibilité et la dynamique des filtres récurrents simulant les résonateurs en fonction de différents types de signaux (transitions V/V, V/NV, NV/NV, NV/V, parties stationnaires et non stationnaires, voyelles longues, ...). Nous indiquons par V (resp. NV) les parties périodiques ou voisées (resp. non périodiques ou non voisées) du signal ;
- l'estimation de la dynamique et de la contribution de chaque variable.

La faisabilité de l'algorithme de synthèse en langage assembleur implique l'évaluation de la précision et de la durée de calcul de chaque module, l'estimation de la taille du programme et l'étude des problèmes de convergence en

présence des troncatures. A cette fin nous avons développé un programme qui permet de simuler, en virgule fixe, les opérations arithmétiques telles qu'elles se déroulent dans le processeur.

Les simulations en virgule fixe font apparaître plusieurs types d'erreurs :

- les erreurs liées à la mise à l'échelle, elles sont indépendantes du processeur,
- les erreurs de quantification lors des conversions A/N et N/A,
- les erreurs de troncature dépendant de l'architecture du processeur,
- les erreurs de débordement,
- les erreurs liées à l'approximation des fonctions mathématiques.

Cependant, les erreurs de quantification et de troncature sont généralement les plus influentes sur la précision. Le meilleur compromis entre la précision des calculs et la garantie de la stabilité a été déterminé par approximations successives. En effet, il est évident que les phénomènes de

saturation et de précision sont des exigences antagonistes qui imposent un compromis.

Une solution pour dégager un compromis saturation/précision exige une attention particulière pour la mise à l'échelle et le choix de la dynamique de toutes les variables.

- Les procédures adoptées pour réaliser ce compromis sont :
- un décalage des variables avant une opération provoquant un débordement,
 - l'augmentation du nombre de bits alloués à certaines variables,
 - l'utilisation de l'unité de saturation arithmétique du processeur.

Le concept de récursivité mis en œuvre pour la réalisation des filtres numériques rend la dynamique de certaines variables intermédiaires dépendant des signaux à traiter. Un corpus de référence, représentant les grandes classes de signaux de parole (segments stationnaires et non stationnaires, occlusions, voyelles longues, gémination, transitions V/V, V/NV, NV/NV, NV/V, ...) est utilisé pour tester la robustesse du système.

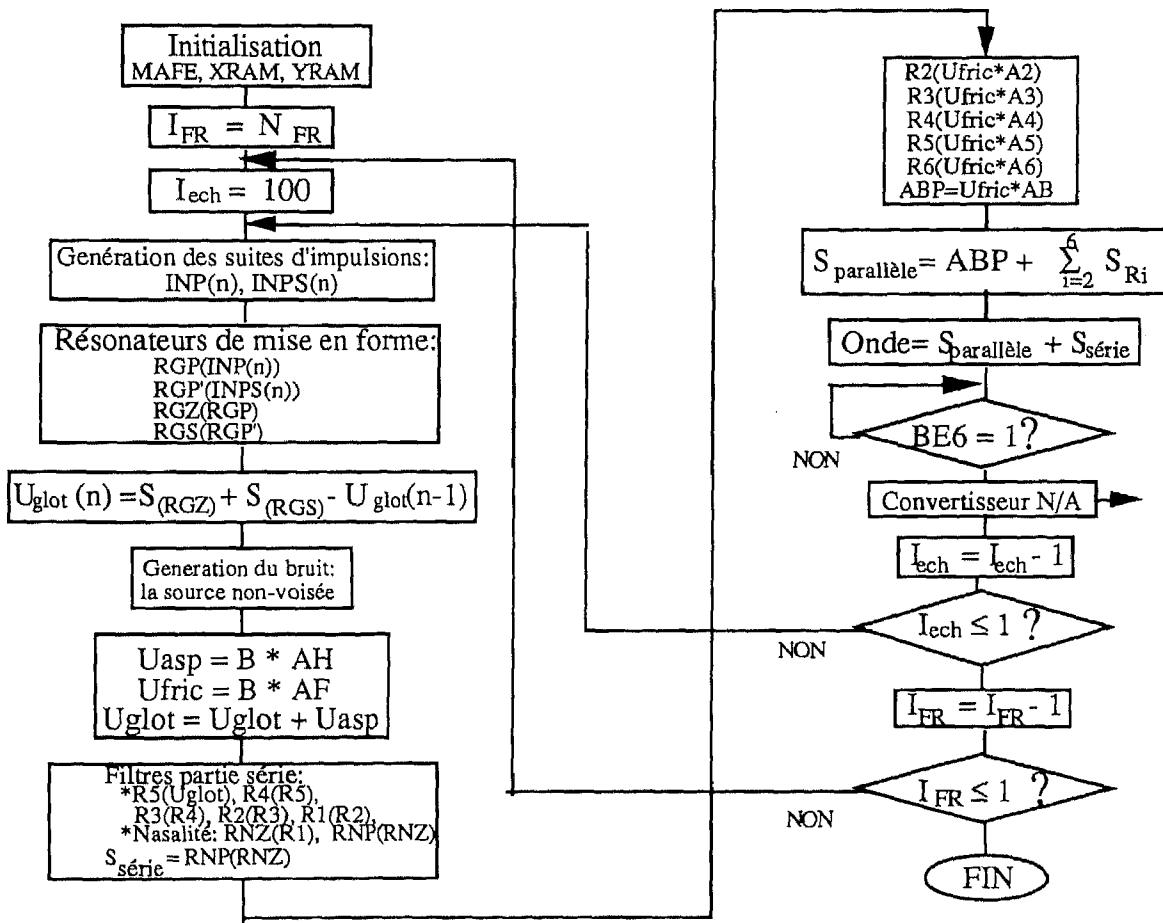


Fig. 2. — Organigramme de synthèse.

3. Implantation

La programmation en langage assembleur englobe quatre étapes :

- la traduction de l'algorithme en microcode du ST18940,
- la simulation en langage PSMMUL,
- l'optimisation de la taille du programme et des données requises,
- les tests en temps réel en utilisant la carte d'évaluation ST18940-EMU.

Bien que ces étapes exploitent certains résultats obtenus en simulation en virgule fixe, elles restent toujours difficiles à réaliser à cause de la taille limitée des mémoires internes. L'opérateur doit gérer de manière optimale le « pipeline » et maîtriser l'exploitation du parallélisme et des codes des boucles. La longueur du mot utilisée dans chaque module de l'algorithme a été choisie après des séries intensives de simulations où nous avons essayé systématiquement d'évaluer l'impact de chaque paramètre sur la qualité de la parole de synthèse. La longueur du mot en mode réel est de 16 bits. Cependant, il s'est avéré nécessaire, pour certains modules, d'utiliser l'arithmétique double précision (32 bits). Tel est le cas pour la partie

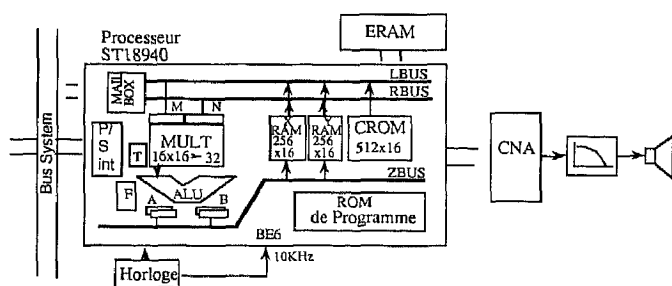


Fig. 3. — Synoptique de système de synthèse.

cascade du synthétiseur (fig. 1) où il s'agit de surmonter (ou minimiser) les problèmes associés aux cycles limites dans les filtres numériques récurrents.

Le bloc-diagramme de l'algorithme est présenté sur la figure 2. La figure 3 représente le schéma de la maquette-prototype. Nous nous sommes imposés comme contrainte majeure la modularité afin de garantir la transportabilité du programme et son adaptation facile aux futures versions du processeur.

Lors du déroulement de l'algorithme en temps réel, les coefficients sont mis à jour au début de chaque trame (10 ms). La fréquence d'échantillonnage délivrée par une horloge externe est égale à 10 kHz.

Les filtres numériques évaluent un échantillon toutes les 100 μ s, lequel est transmis au convertisseur numérique analogique externe de 12 bits. Un filtre passe-bas analogique est connecté à la sortie du convertisseur.

La taille du programme optimisé est de 720 instructions. Le tableau 1 résume les temps d'exécution de chaque module du programme. Le total de 49,9 μ s doit être

majoré de 20 μ s pour les instructions d'initialisation, de test et de branchement. Cette valeur résultante nous donne la possibilité d'augmenter la fréquence d'échantillonnage jusqu'à 14,3 kHz ou d'intégrer d'autres fonctions supplémentaires.

Remerciements

Les auteurs tiennent à remercier le Ministère Français des Affaires Etrangères et le COPEP pour avoir apporté le financement et le soutien qui ont permis la mise en œuvre sur DSP de ce type de synthétiseur dans le cadre de l'Action Intégrée 190/4 LEESA/ENSERB. Ils tiennent également à remercier les lecteurs anonymes et Mr R. Carré pour leurs critiques qui ont permis d'améliorer la présentation et le contenu de cet article.

Manuscrit reçu le 30 juin 1991.

BIBLIOGRAPHIE

- [1] M. C. RAHIER *et al.*, A 3 μ m-NMOS High Performance LPC Speech Synthesizer chip, IEEE J. SSC, vol. SC-18, 3, 1983, p. 349-359.
- [2] F. TANAKA *et al.*, C²MOS Speech Synthesis Systems, IEEE J. SSC, vol. 18, 1, 1983, p. 81-86.
- [3] M. OUADOU *et al.*, Arabic Text To Speech : Single Board, Proc. European Conference on Speech Technology, vol. 2, Edinburgh, 1987, p. 83-86.
- [4] M. OUADOU *et al.*, A Multilanguage Speech Synthesis System, Proc. Speech'88, 7th FASE Symposium, vol. 1, Edinburgh, 1988.
- [5] Y. MASSÉ, G. BENBASSAT, M. NAJIM, M. OUADOU, A CMOS, Synthesizer based on TMS50C42, Proc. EUSIPCO-88, vol. 2, Grenoble, 1988, p. 611-614.
- [6] B. BARAZESH *et al.*, A VLSI Signal Processor with Complex Arithmetic Capability, IEEE Trans. on Circuits and Systems, vol. 35, 1988, p. 495-505.
- [7] M. OUADOU *et al.*, Implementation of 2 400 bps voiceband data modem using the DSP TS68930, 12^e Colloque du GRETSI, Juan-Les-Pins, vol. 2, 1989, p. 845-848.
- [8] A. RAJOUANI, Contribution à la réalisation d'un système de synthèse à partir du texte pour l'Arabe, Doctorat d'Etat ès-Sciences, Université Mohamed V, Rabat, 1989.
- [9] A. RAJOUANI, M. NAJIM, D. CHIADMI, M. ZYOUTE, Synthesis by rule of Arabic language, Proc. Speech tech., vol. 1, Edinburgh, 1987, p. 29-32.
- [10] ST18940/41 Users Guide, SGS-Thomson, 1990.
- [11] CALLIOPE, « La parole et son traitement automatique », Collection Masson, 1989, Paris.
- [12] D. H. KLATT, Software for cascade/parallel formant synthesizer, JASA, 67, 1980, p. 971-995.
- [13] J. ALLEN, M. S. HUNNICUTT et D. H. KLATT, From text to speech : the MITalk system, Cambridge University Press, 1987.
- [14] D. O'SHAUGHNESSY, Design of real-time French text-to-speech system, Speech Communication, vol. 3, n° 3, 1984, p. 233-244.
- [15] G. BAILLY *et al.*, A text-to-speech system for French by Formant Synthesis, Speech 87, Proc. Europ. Speech conf. Edinburgh.

- [16] G. RIGOLL, *The DECTalk system for German : a study of the modification for a text-to-speech converter for a foreign language*, Proc. IEEE-ICASSP, 1987, p. 1450-1453.
- [17] J. M. SANTOS *et al.*, *Text-to-speech conversion in Spanish. A Complete rule-based synthesis system*, Proc. IEEE-ICASSP, 1982, p. 1593-1596.
- [18] Q. T. NGOC et B. GUÉRIN, Synthétiseur à formants numérique en temps réel. Caractérisation, Actes GALF, 15^e JEP, 1985, p. 3-6.
- [19] M. VAN LIEFFERINGE, Etude et réalisation d'un Synthétiseur de Parole basé sur la Méthode des Formants, Revue Etidschrift, vol. 99, 1, 1983, p. 7-27.

		Nombre d'Instructions	Durée d'exécution (µs)
Sources d'excitation :			
+ Bruit blanc		12	1,2
+ modulation d'amplitude		10	1
Résonateurs partie série :			
+ Réel (1 cycle)		202	20,2
+ Double précision (2 cycles)		112	22,4
	Total :	314	42,6
Résonateurs partie parallèle		51	5,1
	Total :	387 (499 cycles)	49,9

Tableau 1

Implantation d'un Synthétiseur de Klatt sur le Processeur ST18940

Titre abrégé	: ST18940 de SGS-Thomson.
Type	: ISFK
Description	: programme d'exécution de l'algorithme de Klatt pour la synthèse de la parole à partir du texte.
Langage	: Assembleur du processeur du signal ST18940 de SGS-Thomson.
Date de création	: 23/11/1990.
Mots clés	: synthèse de parole, formants, temps réel, processeur de signal, simulation, implantation.
Matériel	: Processeur du signal ST18940 de SGS-Thomson, Convertisseur numérique/analogique, Filtre passe-bas, Compatible-PC.
Dispositif particulier	: Haut parleur.
Entrée	: texte
Sortie	: sur haut parleur.
Contact	: M. Ouadou, LEESA Faculté des Sciences, BP 1014, Rabat, Maroc, : Tél. : 19 (212) 7 771838/41/42/44/46, : Télécopieur : 19 (212) 7 774261.
Etat	: en cours de test final d'intégration.
Test	: existence d'une procédure de test.
Maintenance	: pas d'intervention.
Exemple d'application	: — systèmes d'annonces dans les gares, les aéroports, etc... : — dans des systèmes d'aide aux handicapés.