SAPHIR-1: SYSTEME MULTILOCUTEUR COMPRENANT LES PHRASES PARLEES CONTINUES

GRIGORY SLUTSKER

L'Institut d'agriculture d'URSS
B/A 8, Balachikha-9,
Moscou 143900, URSS

SERGUEI KRINOV

L'Institut des problèmes à
transmettre l'information
de l'A.S. d'URSS, GSP-4,
Moscou 101447, URSS

RESUME

Avec le SAPHIR-1, on ne doit pas dicter les mots à l'étape d'apprentissage. Ici, on a un programme de transcription automatique qui compose les étalons phonétiques des mots selon leur forme orthographique comprenant des sygnes d'accentuation nécessaires. Le programme peut être adapté à chaque langue nationale. C'est un expert qui fait l'adaptation, et il n'est pas obligatoire qu'il sache programmer. Les connaissances d'expert s'inscrivent déclarativement en forme des règles de transcription des textes à l'aide d'une métalangue inventée spécialement pour le système. Au régime de compréhension automatique des phrases parlées continues, on garantit un codage phonétique adéquat au niveau de segments.

INTRODUCTION

La plupart des systèmes à reconnaître la parole se fonde sur une procédure d'apprentissage, pendant laquelle le système forme les étalons des mots ou des phrases qui seront employés plus tard. Au cours d'apprentissage l'usager doit dicter tous les mots, et le nombre de prononciations de chaque mot varie de 1 à 5 pour les systèmes divers. Selon des donnés des firmes américaines produisant les appareils à reconnaître la parole, on ne peut assurer la reconnaissance que dans le cas où l'on prononce chaque mot pendant l'apprentissage 3 fois ou plus [1]. Selon les donnés, on sait aussi que la plupart des usagers refuse travailler avec le système si le temps d'apprentissage dépasse 30 min. Il est évident que ces difficultés retiennent l'application vaste des appareils à reconnaître la parole.

Bien des investigations des années dernières ont pour but l'élaboration des systèmes multilocuteurs. Le progrès dans le domaine dépend évidemment des possibilités d'accomplir une analyse phonétique de la parole indépendant de locuteur. Ça simplifie essentiellement le problème d'apprentissage, mais ne le résout pas. Car on doit tout de meme dicter préalablement le vocabulaire, ce qui conserve tous les défauts de l'apprentissage (les défauts principaux sont: des fautes phonétiques incontrolées et des bruits acoustiques qui peuvent déformer les étalons).

SAPHIR-1 permet se passer de microphone et de locuteur à l'étape d'apprentissage. Les mots du vocabulaire en forme orthographique munis des signes d'accentuation s'introduisent dans le système et un programme spécial les transforme aux étalons phonétiques. Pour représenter les étalons, on a choisi le niveau des segments phonétiques, ce qui réduit essentiellement le volume du mémoire nécessaire et le temps des calculs à l'étape de reconnaissance, en comparaison avec une représentation paramétrique. En outre, le procès de segmentation normalise le signal parlé dans le temps, car le nombre de segments et nombre de sons du mot sont approximativement les mêmes. La transformation automatique "orthogramme — conséquence de segments" se fait selon les règles reflétant les lois de prononciation de la langue en question. On tient aussi compte des charactéristiques techniques du segmenteur et du marqueur qui effectuent l'analyse acoustico-phonétique dans le système.

La méthode permet représenter le vocabulaire à apprendre en forme de morphèmes. La représentation par éléments est irréalisable aux systèmes exigeant l'enrégistration phonique préalable, car on ne peut pas prononcer certains éléments en isolation correctement (tels sont les morphèmes inaccentués). Le variant du vocabulaire réalisé en SAPHIR-1 présente plusieurs morphèmes inaccentués, par exemple, l'élément ЗАР qui a dans la grammaire le meme droit que le mot ЗА'РАБОТНАЯ en combinaison avec le mot ПЛА'ТА.

Dans le SAPHIR-1, c'est la programmation dynamique qui aide à choisir l'étalon le plus proche au signal d'entrée sur le niveau segmentatif. L'emploi de la programmation dynamique pour la reconnaissance automatique de la parole a été premièrement proposé en 1968 par un des auteurs du rapport [2].

En ce qui concerne les pricipes de l'analyse acoustico-phonétique accomplis dans le SAPHIR-1, on les a vérifiés aux systèmes multilocuteurs à reconnaître les mots isolés [3].

FORMATION AUTOMATIQUE DES ETALONS PHONETIQUES

Créant l'algorithme de transformation "orthogramme → conséquence de segments", on est tombé sur les difficultés suivantes. L'algorithme reflétant toutes les variations phonétiques de la langue en question et toutes les charactéristiques nécesaires de l'analyseur exige un programme au grand nombre d'opérateurs de transfert conditionnel. Les programmes de la sorte sont difficiles à mettre à point. Il est encore plus difficile de changer le programme. En réalité, le besoin de changement peut surgir, par exemple, dans le cas de mise en vue des lois phonétiques supplémentaires ou dans le cas d'emploi d'un analyseur acoustico-phonétique ayant un nouveau alphabet de segments. En outre, l'algorithme reflète l'organisation phonique d'une seule langue nationale.

Ce qui serait le mieux, c'est un algorithme souple qui permet modifier ou meme remplacer les règles de transformation, par exemple, au changement de la langue nationale en question. Pour atteindre ce résultat, on peut s'orienter aux connaissances d'expert prises comme un massif d'entrée pour le traitement des orthogrammes des mots. Un expert-phonéticien introduit ses connaissances en forme d'un ensemble ordonné des règles déclaratives de transformation. Il ne doit pas savoir programmer, car on emploie ici une métalangue adaptée pour l'inscription spectaculaire et l'interprétation automatique des règles. Donc, le programme à transformer les textes c'est un programme "comprenant" la métalangue [4].

Cette méthode augmente la souplesse du système et le programme meme devient universel. On entend l'universalité comme ça.

1. Comme l'alphabet de sortie, on peut prendre tout alphabet nécessaire décrit par les règles de transformation - seraient-ce les codes phoniques, les symboles de la transcription phonétique internationale, les symboles alphabétiques d'une autre langue nationale, etc. On peut employer la possibilité dans les manuels de conversation (la formation automatique de prononciation), pour la translittération, etc. On peut aussi user le programme aux systèmes de synthèse "texte — parole" (dans ce cas, l'alphabet de sortie consiste des codes à diriger le synthétiseur du sygnal parlé).

2. Les règles nécessaires étant données, on peut traiter un texte écrit en langues différentes.

METALANGUE POUR LES REGLES DE TRANSFORMATION

La métalangue à inscrire les règles de transformation suffit pour transformer un texte en forme phonétique et pourtant elle est très simple et évidente. Pour l'employer, on ne doit pas savoir programmer.

L'aspect général d'une règle de transformation est [A_B], où A est symbole ou chaîne de symboles, B est symbole ou chaîne de symboles qui est le résultat de transformation d'A. La chaîne A peut correspondre à un fragment d'orthogramme (fragment en code d'entrée) ou à un fragment d'une représentation intermédiaire. La chaîne B correspond à un fragment d'une représentation intermédiaire ou à un fragment en code de sortie.

Les représentations intermédiaires ne sont qu'auxiliaires; ce sont elles qui aident à donner une classification traditionnelle des symboles - par exemple, une classification selon la mode ou le point d'articulation; à l'étape on peut aussi introduire les marques spéciales pour les voyelles pré- et post-toniques, etc.Pour l'expert-phonéticien, la représentation intermédiaire de la sorte est facile à interprendre, car elle reflète les lois phonétiques de la langue en question. En cas de nécessité, les règles de transformation forment une image phonétique très détaillée - on peut p.e. décrire le charactère supposé des trajets de formants aux frontières "consonne-voyelle", etc. Dans le SAPHIR-1, ayant un analyseur acoustico-phonétique à un 16-segment-alphabet, les possibilités de la représentation intermédiaire sont limitées.

Les chaînes A et B des règles sont séparées par "_"; les règles sont toujours mis aux crochets. Tout ce qui est outre les crochets n'est qu'un commentaire. Les commentaires font le système des règles plus évident; l'ordre des commentaires n'est pas fixe. La mode de la mise en ordre des règles peut être choisi librement. P. e., on peut les donner en forme d'une table dont les lignes et les colonnes unissent les règles selon tels ou tels indices; les titres des lignes et des colonnes peuvent présenter les commentaires - comme les symboles des processus d'assimilation (d'assourdissement, d'avoisement, etc.).

Pour qu'on puisse écrire les règles en forme générale tenant compte d'analogie propre à certaines transformations,on a introduit des variables pour nommer les symboles de texte.

N'IMPORTE QUEL SYMBOLE. Le symbole © marque n'importe quel symbole - p.e., la règle [A©B_AB] élimine tout symbole placé

dans le texte entre A et B.

N'IMPORTE QUEL SYMBOLE D'UN ASSORTIMENT. La règle [<ABCDE>F_@Z] remplace le symbole F par le symbole Z si F succède à n'importe quel symbol de l'assortiment mis aux parenthèses angulaires. L'introduction d'une variable (marquée par parenthèses angulaires) permet généraliser les règles analogiques. On peut introduire dans une règle jusqu'à 10 variables; dans ce cas, on leur donne les index de 0 à 9. P.e., la règle [<AB>T<CDE>1F_@1@F] décrit les transformations éliminant le symbole T et transposant les symboles qui l'entourent.

N'IMPORTE QUEL SYMBOLE À L'EXCEPTION DES SYMBOLES NOMMÉS. Une variable de la sorte peut être désignée par ٦...> ; le signe recoit le nom des parenthèses avec la negation. La règle transforme le contexte sous la condition d'absence des symboles entre parenthèses avec la négation - p.e., [B٦UVW>_ZF@] ou [RA٦UVW>_RAW@] . Fixée par le contexte, la signification d' une variable peut faire partie d'une autre variable:[<JLNZ>@1<JLNZ>2?@1>3_@@1w@2@3].Les types des règles démontrés au-dessus permettent décrire toute transformation du texte en n'importte quelle langue.

CLASSIFICATION DES SONS DE LA PAROLE.

Pour qu'on puisse résoudre le problème de transformation ORTHOGRAMME →DESIGNATIONS PHONETIQUES, on doit paramétriser adéquatement la description des sons de la parole correspondant aux lettres. Tenant compte des possibilités de l'analyseur du SAPHIR-1, on a choisi une application de deux dimensions convertissant des symboles orthographiques à une représentation intermediaire.

Tous les symboles désignant les voyelles comprennent soit U (voyelle tonique) soit V (voyelle première prétonique ou atonique), soit W (voyelle réduite). Parfois, les réduites n'ont pas de correspondance avec les symboles orthographiques - ce sont d'habitude les intercalations réduites entre les consonnantes voisines.

Si la transcription automatique exige qu'on donne une charactéristique plus détaillée de la dépendence existant entre l' accent et les qualités des voyelles, on peut le faire à l'aide de la méthode proposée. On doit introduire les désignations pour les types des sons en question et composer les règles nécessaires.

Les lettres correspondant aux consonnes, on les remplace par une couple de symbole dont le premier signifie un des huit modes d'articulation possibles. L'autre montre un des quatre points d'articulation propres aux consonnes. Le codage de la sorte permet donner les règles dans une forme généralisée et évidente.

Dans la table ci-dessous, on peut voir les transformation faites srlon la méthode pour les consonnes russes. Les symboles de la table ont les significations

	F	D	A	*	G
Q	[П_QF]	[Т_QD]			[К_QG]
J	[Б_JF]	[Д_JD]			[Г_JG]
L	[В_LF]		[Л_LA]	[Й_L*]	
N	[М_NF]	[Н_ND]			
R			[Р_RA]		
Z		[З_ZD]	[Ж_ZA]		
S	[Ф_SF]	[С_SD]	[Ш_SA]	[Щ_S*]	[Х_SG]
C		[Ц_CD]		[Ч_C*]	

suivantes: Q - les occlusives sourdes, J - les occlusives voisées, L - les liquides, N - les nasales, R - les vibrantes, Z - les constrictives voisées, S - les constrictives sourdes, C - les affriquées;

F - les labiales, D - les dentales, A - les alvéolaires dures, * - les alvéolaires molles, G - les postpalatales.

Le groupe des règles, y compris les titres de la table, se trouve dans le massif d' entrée du programme de transformation. Le programme prend les titres comme commentaires et ne travaille pas. La table montre que les lettres russes correspondant aux consonnes n'épuisent pas toutes les positions. Une autre langue nationale peut donner une autre disposition des symboles dans la table par exemple, la " " ucrainienne doit être présentée comme "ZG" Pour d'autres langues, on peut proposer des tables variant le nombre des lignes et des colonnes.

Maintenant on peut citer les règles d'assourdissement et d'avoisement - par exemple, [J@<;;:. >1_Q@@1] (l'assourdissement des voisées à la fin du mot); [S@<JZ>1_Z@@1] (l'avoisement des constrictives sourdes devant les voisées - occlusives et constrictives).

MISE EN ORDRE DES REGLES TRANSFORMATIVES.

Toutes les règles de l'assortiment sont mises en ordre; on peut les diviser en quatre groupes.

Le premier groupe consiste des règles générales qui agissent tout d'abord,avant la transition à la représentation intermédiaire de deux dimensions, et des règles transformatives pour les voyelles - par exemple,[<шцж>И_@ы] (le remplacement d' "И" par "ы" après "ш","ц","ж" en russe).

Le seconde groupe comporte les transformations des consonnes; la plupart des règles est présentée dans la table. En outre, dans le groupe entrent les règles d'avoisement, d'assourdissement et de fusion des consonnes, les règles d'intercalation des réduites entre les consonnes voisées et d'intercalation des sourdes aspirées entre les consonnes sourdes, etc. Ce sont 80 règles environ qui font partie des deux groupes. Pour la langue en question, un expert-phonéticien introduit les règles à la fois: on n'a pas besoin de renouvellement de l'information.

Le troisième groupe permet la transition aux codes des segments phonétiques, qui peuvent être fixés par le segmenteur automatique du système.

Enfin, c'est le quatrième groupe des règles qui contrôle le fonctionnement du programme transformatif. Les règles font la transformation inverse qui donne à la sortie les orthogrammes initiales. On n'emploie pas les règles à former les étalons phonétiques des mots - elles ne servent qu'à contrôler.

D'AUTRES NIVEAUX DU SAPHIR-1.

Le SAPHIR-1 a 10 niveaux hiérarchiques: celui acoustique, celui paramétrique, segmento-phonétique, orthographo-morphologique, lexique, sémantique, syntaxique, phraséologique, dialogique, pragmatique. Donc, le niveau acoustique et celui pragmatique sont deux bouts inverses du système dont le principe de fonctionnément est suivant.

Etant donné les connaissances sur tous les niveaux obtenues à l'aide des experts-spécialistes, le SAPHIR-1 est prêt à comprendre les phrases en chaque situation de dialogue homme-machine avec n'importe quel locuteur.

D'une part, chaque phrase prononcée par celui-ci passe à travers les trois premiers niveaux de traitement et le signal parlé se transforme en une suite des segments phonétiques. Simultanément de son autre bout, le système engendre généralement parlant toutes les phrases (en meme forme phonétique) qui correspondent à la situation du dialogue.

Deux mouvements inverses se rencontrent au niveau phonétique à comparer les successions de segments. La comparaison s'effectue rapidement selon le graphe syntaxique à l'aide de la méthode DP [2] modifiée à reconnaitre la parole continue. Ainsi, le SAPHIR-1 trouve une phrase engendrée la plus proche à celle prononcée.

Au niveau lexique, chaque orthogramme s'accompagne par un code sémantique comprenant 4 symboles par correspondance au domaine thématique en question, la classe de notion, l'objet concret de la classe et les charactéristiques grammaticales de la forme. De cette façon, le vocabulaire du système consiste des couples "orthogramme-code sémantique". Cela permet différer entre eux les HOMONYMES coïncidants phonétiquement mais qui se distinguent par leurs codes sémantiques. Et au contraire, les SYNONYMES ont presque même code sémantique mais ils ont des formes divers.

Au niveau syntaxique à chaque situation du dialogue, l'usager peut exploiter un sous-ensemble des phrases qui est déterminé par la grammaire situative correspondante. Celle-ci est représentée par un graphe. Le noeud de sortie de celui-ci tient toutes les variantes des phrases de la grammaire. Après la fin du résonnement d'une phrase prononcée par l'usager, le SAPHIR-1 prend le résultat final au noeud de sortie et trouve la trajectoire la plus vraisemblable le long du graphe à reconstruire la suite des mots. Ce sont les orthogrammes des formes que l'on use pour l'extraction du texte à l'écran (ou pour la synthétisation de la parole) et la succession des codes sémantiques à calculer le sens de phrase.

Le Système Automatique à comprendre des PHrases contInues paRlées (SAPHIR-1) etait réalisé à la base de l'ordinateur du type PDP 11/40 [5] .

La sureté de reconnaissance de la parole continue pour n'importe quel locuteur n'est pas moins que 95%. Le vocabulaire d'exemple consiste de 200 formes. Son volume n'est pas restreint par les possibilités du système.

BIBLIOGRAPHIE.

1. Meng B. Speech recognition: not a typical engeneering problem. Digit Des., 1985, n6, p. 49 - 57.
2. Слуцкер Г.С. Нелинейный метод анализа речевых сигналов. Труды Научно-исследовательского института Радио (НИИР),М.,I968г. вып.2. с. 76 - 82.
3. Кринов С.Н., Савельев В.П., Цемель Г.И. Классификация сегментов при распознавании устных команд. Труды АРСО-I3,Новосибирск, Институт математики СО АН СССР,I984,с.IOI-IO3.
4. Кринов С.Н., Слуцкер Г.С. Автоматическое формирование звуковых эталонов слов по их орфографической записи. Труды АРСО-I4, ч.I, Каунас, I986, с.86.
5. Кринов С.Н., Слуцкер Г.С. Многоуровневая речевая диалоговая система "САПФИР". ТРУДЫ АРСО-I4, ч.I., Каунас, I986., с.92-94.