SPEECH SYNTHESIS FROM TEXT BASED ON SYLLABLES

T. Radhakrishnan and R. Castillo

Computer Science Department
Concordia University, Montreal

ABSTRACT

Words, morphs, syllables, diphones, phonemes and microphonemes have been considered by different authors as concatenative units for speech synthesis. Generation of speech from written texts is becoming more and more important due to the diversity of its applications. In this paper, we give a brief introduction to the orthography, phonetics and phonology of Spanish. Then, we present an algorithm to transform a written sentence in Spanish into a sequence of syllables along with stress markers, pause markers, and intonation-level markers. These markers and the syllables would form an input to a voice synthesizer unit. Also, we describe the application of this technique to sentences generated by a tranformational generative grammar that can be interfaced to a data base.

RÉSUMÉ

Les mots, les morphs, les syllabes, les "diphones", les phonèmes et les microphonèmes ont été considérés par différents auteurs comme des unités d'enchaînement pour la synthèse de parole. La génération de la parole à partir de textes écrits, devient de plus en plus importante à cause de la diversité de ses applications. Dans cet article, une brève introduction est donnée sur l'orthographe, la phonétique et la phonologie de l'espagnol. Ensuite, un algorithme est présenté pour transformer une phrase écrite dans cette langue, en une séquence de syllabes ainsi qu'avec des indicateurs d'accent, des indicateurs de pause et des indicateurs de niveau-d'intonation. Ces indicateurs et ces syllabes pourraient former des données d'entrée à un synthétiseur de parole. On décrit aussi, l'application de ces techniques à des phrases produites par une grammaire génératrice transformationnelle qui pourrait être liés à une base de donnée.

I.-INTRODUCTION:

Speech synthesis has a relatively long history and yet is still an important topic in speech research[1]. Developments in digital electronics, microprocessors, and the interfaces between computers and communication equipment, all have added more importance to speech synthesis. Methods of speech synthesis are based essentially on models of speech production[2]. Such models may be focused at different levels of speech production, such as semantics of the discourse, syntax of the language, phonology, motor control, articulatory movement, and sound formation. Two major approaches to speech synthesis are: a) synthesis by analysis, and b) synthesis by rule. In synthesis by analysis, the control parameters of speech are derived directly from an analysis of humanly produced utterances. These parameters are stored, retrieved upon demand and used for synthesizing the corresponding utterances. The goal of synthesis by rule is to generate natural sounding speech from symbolic input by means of rules[3].

For speech synthesis, words, morphs, syllables, diphones, phonemes and microphonemes have been considered as concatenative units by different researchers. In voice response systems that use a limited vocabulary, perhaps a word based approach is adequate. It had been observed that about 200 kilo bytes of disc storage would be adequate to store approximately 1000 words, with an average duration of 2/3 seconds per word, when adaptive DPCM method of coding is used[4]. Many English words have internal structure, made up of atomic units called "morphs"[5]. The morphs include suffixes, prefixes, and roots. On the average, there are about two morphs per English word. It is found that the number of morphs in English (approximately 12,000) is at least an order of magnitude less than the number of words in the language. Hence, for synthesis of speech from unrestricted texts, morphs are more attractive than words.

One or more syllables together make a word in a spoken language. From the articulation point of view, the utterance of a syllable is a cyclic process that passes from onset to peak and peak to offset as the vocal tract moves from a more-closed to a more-open to a more-closed configuration[6]. Syllabication of words is one of the problems in the synthesis of speech from texts using syllables. A great deal of research based on "syllables" is in progress at Haskins Laboratories[6], Bell Laboratories[7], Bell-Northern Research[8], and at other institutions[9].

In [10], Bertinetto et al., use diphones or pairs of phonemes, as concatenative units for the synthesis of speech from written Italian texts. They have used about 150 diphones to generate the required sound in Italian. Today, voice synthesizer units and LSI chips for phoneme based synthesis are commercially available, even for personal computer users. Votrax is one of the popular commercial voice synthesizer units based on phonemes[11]. Methods for automatic translation of English text to phonetics by means of letter-to-sound rules have been discussed by Elovitz et al. in [12], and by McIlroy in [13]. For example, McIlroy's system contains more than 750 letter-to-sound rules which include 100 words, 580 word fragments, and 70 letters. The system also has a small 100 word exception dictionary. On the other hand the NRL system is driven by a set of 350 letter-to-sound rules. Both systems produce phonetic transcriptions from the input text.

For synthesizing speech from Polish texts, Kielczewski had used concatenative units smaller than phonemes which he called as "microphonemes"[14]. According to him, microphonemes are short characteristic segments of audio signals(40 to 60 msec) which after certain number of repetitions(1 to 20, depending on the type) enable correct perception of the phonemes they represent. He observes that the microphonemes of fricative phonemes are longer and that of voiceless stop consonants are a little shorter. By lengthening the duration of microphonemes, and changing the number of repetitions and the signal amplitude, he is able to produce the effect of stress and intonation.

The following factors are to be considered in the choice of concatenative units(CU) for synthesizing speech from written texts:

a) The necessary modifications and preprocessing of the input text to suit the algorithm (set of rules) under consideration.
b) Rules for the transcription of an input text into a sequence of concatenative units.
c) Transformations within CUs and between adjacent CUs when the units are concatenated to produce synthetic speech.
d) The complexities involved in the placement, and the realization of the effects of prosodic features so as to make the synthetic speech natural sounding.
e) Synthesizer units to suit the needs of the CUs.

In the foregoing paragraphs we referred to different concatenative units. The transformations in those CUs when conacatenated, have been studied extensively by different researchers. For example, the tranformations of microphonemes when they are concatenated, have been reported in [14], that of phonemes are discussed in [15], and the problems associated with syllables and syllable boundaries are considered in [7].

Synthesis of speech from texts writeen in different languages have been studied in the literature. In [16], for example, texts written in English, Russian, Spanish, Italian, and Esperanto have been considered. Texts written in Polish have been studied in [14] and text written in French have been considered by different researchers[17]. The underlying approach of a text-to-speech procedure may be the same when applied to different languages, but there are differences in the surface level structures from one language to another. In this regard, our discussions in the following sections are mostly pertinent to the Spanish language.

II.-ON SPANISH ORTHOGRAPHY AND PHONOLOGY:

Orthography is a set of rules to describe a spoken language in written form. The relationships between the phonological and the graphic alphabets are generally established by a set of rules. There are exceptions to such rules. In the Spanish spelling system there is a close correspondence between the strings of graphemes and the sound(s) they stand for. The Roman alphabets A to Z and a to z form the base for the graphemes used in the Spanish spelling system. In addition to this, there are three more graphemes (ch, ll, ñ) and two diacritical marks known as orthographic accent(') and dieresis(¨). The accent can be associated only with the vowels(a, e, i, o, u). The dieresis is used only with u and that too in some of the cases where it is preceded by g and followed by e or i[19]. The inverted question mark and the inverted exclamation mark are used in the begining of an interrogative and exclamatory sentences respectively. The other punctuation marks , ; . : ! and ? are used in a similar way as in English.

According to Navarro[21], there are 42 phonemes in the Spanish phonological inventory. There are five phonemes corresponding to the vowels which give rise to twenty allophones. Similarly, the nineteen phonemes that correspond to the consonants give rise to over thirty allophones. These nineteen phonemes are: /b/, /č/, /d/, /f/, /g/, /j/, /k/, /l/, /ḽ/, /m/, /n/, /ñ/, /p/, /r/, /r̄/, /s/, /t/, /y/, /z/. There are six falling diphthongs /ai/, /au/, /ei/, /eu/, /oi/, /ou/; eight rising diphthongs /ia/, /ie/, /io/, /ua/, /ue/, /uo/, /iu/, /ui/; and four triphthongs /iai/, /iei/, /uai/, /uei/. Although from the phonetic point of view diphthongs and triphthongs can be divided into vowels, semivowels, and semiconsonants, in Phonology they play the same role as single phonemes[21]. Some researchers disagree with this.

In general the mapping between

graphemes and phonemes is many-to-many. For example, the phoneme /b/ is associated with the graphemes b and v; /k/ is associated with the graphemes c, qu, or k. The grapheme h has no associated phoneme. Depending on the context, the grapheme x may be associated with any of the phonemes /k/+/s/, /s/, /j/. For our experiments in text-to-speech, we have used the letter-to-sound rules shown[19,20] below:

```
U = {a,b,...,y,z}, Y∈U, A∈{a,e,i,o,u}
B∈{a,o,u}, C∈{b,c,ch,...,y,z}, E∈{e,i}
D∈U - {e,i}, # is a word boundary
guE-->/g/E, guB-->/gu/B, guY-->/gu/Y
 gE-->/j/E,  cE-->/z/E,   cD-->/k/D
quE-->/k/E,  xC-->/s/C,  AxA-->A/ks/A
 ch-->/č/,   ll-->/ḻ/,     ñ-->/ñ/
 #r-->/r̄/,   rr-->/r̄/      r-->/r/

  b-->/b/, v-->/v/, d-->/d/, f-->/f/,
  j-->/j/, k-->/k/, l-->/l/, m-->/m/,
  n-->/n/, p-->/p/, s-->/s/, t-->/t/
  w-->/g/, y-->/y/, z-->/z/, a-->/a/,
  e-->/e/, i-->/i/, o-->/o/, u-->/u/
```

The individual sounds of vowels, consonants, and diphthongs are commonly referred to as segmental phonemes. In spoken utterances, they are influenced by suprasegmental features such as stress, pitch, rhythm, and intonation. The necessary information for the placement of suprasegmental features are mostly embedded in a text. Stress, the prominent syllable within a word, can be predicted for Spanish text by means of the following rules:

R1 Words ending in a vowel sound, /n/, or /s/ tend to be stressed on the next to the last syllable.

R2 Words ending in a consonant sound other than /n/, or /s/ tend to be stressed on the final syllable.

R3 An orthographic accent when present will take precedence over R1 and R2.

In the case of Spanish, it is very important to place the stress on the right syllable in a word. In many cases, the placement of stress distinguishes between present and past tense(estudio, estudió); between a noun and a verb(término, terminó); or between adjective and verb(corto, cortó). Similarly, intonation enables a speaker/listener to differentiate among affirmative, interrogative and exclamative sentences. Pauses are used in a speech either for emphasis or for breathing(air intake). It has been estimated that in Spanish about eight to ten syllables constitute a "phonic group" or the speech segment between two consecutive pauses meant for breathing[19, 20, 21]. However, smaller and larger phonic groups are not totally absent.

There have been several attemps to define the term "syllable", but still there is no commonly agreed definition[9, 23]. Also there are variations in the interpretation of syllables from language to language. A syllable in Spanish always has a vowel sound as its nucleus which may or may not contain additional consonant sounds[20].

III.-SPEECH SYNTHESIS FROM WRITTEN TEXTS:

Experimental systems for the synthesis of speech from texts written in Spanish have been considered by Berdichevsky et al. in [24], and Sherwood in [16]. The Computalker CT-1 synthesizer and the Votrax VS-6 have been used respectively in the above two experiments. In both cases phonemes have been considered as concatenative units.

The parameters in Votrax VS-6 are tuned to synthesize General American English (GAE). This introduces limitations when a different language like Spanish is synthesized using Votrax VS-6. In particular, the phoneme /r̄/ (alveolar trill) of Spanish is absent. Substitution like /r/ [16] will cause confusion between words like pero(but) and perro(dog), enterar(to inform) and enterrar(to bury). Other phones like those used to produce the GAE phonemes /p/, /t/, /k/ are aspirated but in Spanish they are not and this difference will cause confusion between words like: tomar(to take) and domar(to tame). Depending on the particular application, Votrax VS-6 might be practical for synthesizing Spanish or might present serious limitations. Multilingual versions of Votrax are now available.

Motivated by the following factors, we have selected syllables as the primary concatenative unit for our experiments.

a) Although there are some exceptions[19], it is possible to syllabicate the words in a Spanish text rather easily.
b) Syllables tend to be more or less of the same duration[22, 23].
c) Syllables being higher order concatenative units than phonemes, for a given text, the number inter unit boundaries will be small when syllables are used.
d) Spanish is a syllable-timed language that is the syllable is the basic rhythm unit of Spanish[20, 23].
e) Placement of stress on a syllable within a word and formation of intonation contours over a sequence of words are relatively easier in Spanish when syllables are considered.

The syllable structure of Spanish has been studied by different linguists. In particular Navarro studied the syllabic types in Spanish texts[21]. Delattre[22] compared the syllable frequencies and syllable structures of different languages with that of Spanish. The syllabic type CV has the highest frequency of occurrence, almost 60%[21], and there are 95 such CV pairs. However, the number of different syllables is very large, when compared to the number of phonemes. The combined approach of syllables and phonemes is not ruled out at least theoretically.

In order to specify the intonation contours for non-emphatic spoken utterances, we use three intonation levels. The spoken utterance of a sentence or part of a sentence starts at level one and shifts to level two at the occurrence of the first stressed syllable and remains there until the occurrence of the final stressed syllable. For affirmative sentences, the level finally drops to one while for interrogative sentences the intonation level will rise to three. Also in certain non-interrogative sentences, such as "si lo quieres hacer"(do it if you want) the level rises to three. We use the term "orthographic-phrase" to denote a sentence or part of a sentence that is terminated with a pause while reading. In our case, an orthographic phrase is delimited by punctuation marks. The following procedure is proposed for synthesis of speech from text using syllables. Its goal is to extract as much information as possible from the orthographic text and then to "speak out the output buffer contents".

```
PROCEDURE READ-SPEAK

1 Get an orthographic phrase;

2 Get a token from the orthographic
  phrase;

3 Repeat steps 4 thru 9 till the end of
  the data is reached;

4 Repeat steps 5 thru 6 till the end of
  the orthographic phrase is reached;

5 Case TOKEN of
  WORD :
      Generate phonetic transcription
      of the word,
      Syllabicate the phonetic word,
      Predict the syllables to be
      stressed placing stress markers,
      and
      Place the result onto the output
      buffer;
  PUNCTUATION MARK :
      Place pause marker in the output
      buffer;
  EXCLAMATION OR QUESTION :
      Place sentence type marker in the
      output buffer
  End Case TOKEN of;

6 If there are more than 8 syllables in
  the phonic group then add a pause in
  the output buffer to indicate a
  breath group; .

7 Get the next TOKEN from the
  orthographic phrase;

8 Based on the sentence type, introduce
  intonation level markers in the
  output buffer,
  Speak out the buffer contents;

9 Get the next orthographic phrase;
End PROCEDURE READ-SPEAK
```

The syllabication referred to in step 5 of the above procedure makes use of the following rules that are derived from the linguistic rules found in [19, 20]:

Let C denote a consonant sound, V a vowel sound, "-" a syllable boundary, and the set P be defined as:

$\{C_1C_2 \mid (C_1 \in (/p/, /b/, /f/, /g/, /t/, /k/)$ and $C_2 \in (/r/, /l/))$ OR $(C_1C_2 = /dr/)\}$

S1 The sequence VCV is divided as V-CV

S2 The sequence VC_1C_2V is divided as

a) $V-C_1C_2V$ if $C_1C_2 \in P$

b) VC_1-C_2V otherwise

S3 The sequence $VC_1C_2C_3V$ is divided as

a) $VC_1-C_2C_3V$ if $C_2C_3 \in P$

b) $VC_1C_2-C_3V$ otherwise

S4 The sequence V_1V_2 is divided as V_1-V_2 only if either V_1 or V_2 is a stressed i or u; otherwise the V_1V_2 pair is not separated.

S5 Three vowels in a sequence are never separated

The above rules have been programmed and tested with a set of the most frequent Spanish words[18]. Based on this study, below we list some of the most frequent syllables: /de/, /el/, /la/, /a/, /ke/ ,/i/, /en/, /no/, /e/, /to/, /yo/, /do/, /es/, /por/, /su/, /mo/, /o/, /na/, /ko/, /te/, /un/, /kon/, /ra/, /u/, /pa/, /lo/

For our study in speech synthesis based on syllables, we have created a dictionary of 60 most frequent syllables as follows: the syllables have been spoken by two speakers, under "noise free" conditions and recorded using a good quality tape recorder. The time domain signal is then filtered at 4.5KHz and digitized at a 10KHz rate with 12 bit resolution per sample. The digitized samples are stored on disk for further analyses. We have obtained the spectrograms of the syllables using the Kay Spectrograph for studying their formant transitions. One of our interests is to examine the techniques for smoothing the intersyllable boundaries and for varying the intra syllable parameters in order to produce the effects of suprasegmental features. However, our immediate interest is in the use of speech for man-machine communication in database applications.

IV.-AN APPLICATION TO USER INTERFACE WITH DATABASES:

There is a growing trend in the use of database concepts for storage, retrieval, and maintenance of large volumes of corporate data. Many people who are not experts in computer programming have a need to access these databases for different purposes. Here one of the difficulties lies in the man-machine interface between such naive users and the database management software. For a long time researchers in the area of artificial intelligence have been interested in the development of a natural language understanding system for the communication from a man to a machine[25]. Much work has been done also in the more specialized area of natural language input to databases[26,27]. In [28], Fallside and Young consider the problem of "sentence generation" for the natural language output from a database. We shall refer to these two problems as "natural language understanding"(NLU) intended for database input, and sentence or text synthesis(TS) meant for database output.

Clearly NLU can be combined with speech recognition and TS with speech synthesis to provide speech communication between the database users and a database. As noted by Hill[29], "..speech offers humans a means of spontaneous, convenient and effective communication for which neither preparation nor tools are required, and which may be adapted to suit the developing requirements of a communication situation..". We believe that speech is one of the well suited channels for the communication between a database and its naive users. Both NLU and unconstrained speech recognition are more difficult problems than TS and the synthesis of natural sounding speech. Let R be a sentence produced as a response to a query Q. Which words in R should receive stress would depend on Q. A detailed example of this can be found in [29]. Also, this approach has been applied by Fallside and Young[28] to a very specialized case of producing speech output from a computer controlled water supply network. In a database context, the necessary knowledge for the placement of suprasegmental features are available from two sources: a) in the dialogue of an interactive session between man and machine, b) in the "data dictionary"[30] that contains data (or semantics) about the data stored in the database.

In interactive applications, a speech output generated by the computer could be meant for one or more of the following purposes:

a) to request the user for completion of an incompletely specified query or a command;
b) to report to the user about a syntax error, run time error, spelling error, etc.;
c) as the restatement of a user query for confirmation; or
d) to present the retrieved output for a user query.

The work reported in [28] is mainly concerned with d) above. For synthesizing voice output, we could use the retrieved output for a query along with a data dictionary or a knowledge source, and a grammar such as augmented transition network grammar [31] or transformational generative grammar [32]. It is possible that a spoken output of the sentence, such as, "there is no supplier for this part in Montreal but there are two in Toronto" could be the response from a database to a user query. Based on the user's query to which this sentence is a response, stress could be placed on an appropriate word in the above sentence. Had the question been "Find a supplier for the part X234, preferably in Montreal", we might stress the words "no" and "but" in the above response.

In the following paragraphs, we present a simple case study based on students registration in courses. Simplifying assumptions have been made to make the presentation brief.

<u>Relational schema of the sample database</u>

STUD-BIO:
<ID#,name,TEL#,sex> Bio-data

CORS-FIN:
<ID#,COURSE#,GRADE> Courses finished

CORS-REQ:
<ID#,COURSE#> Courses desired for registration

PRE-REQ:
<COURSE#,PRE-REQ#> Prerequisites for a course

CORS-OFR:
<COURSE#,SECTION#, time, place, session, prof#, capac-rem>

REGI-STA:
<COURSE#,SECTION#,ID#>
Registered students

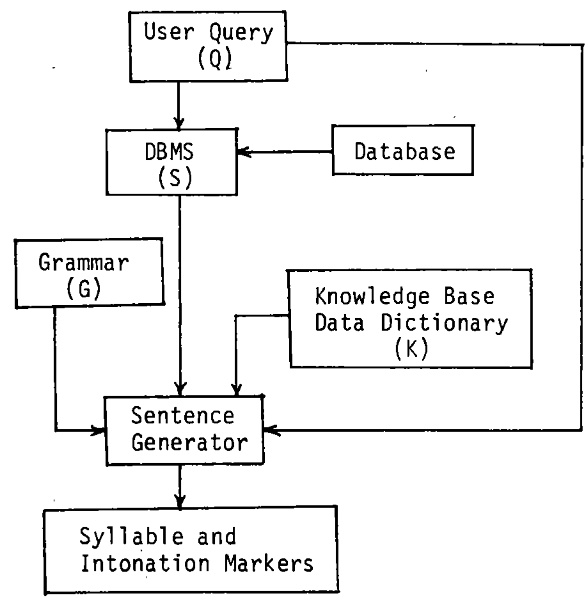

Fig. 1 Block Diagram for Speech Output.

The various stages of a system intended for speech output to the database users are shown in Fig.1.

Two sample queries(expressed by the users, perhaps in a SEQUEL like language):

Q1 How many students did pass the

course COMP352?

Q2 Which professor teaches the course COMP352?

Suppose the responses to Q1 and Q2 are:
a) Thirty students passed the course COMP352.

The system shown in Fig. 1 has generated an affirmative sentence P1 as given below.

S --> NP VP
NP --> det N
det --> $[[[[\text{treinta}]_{ord}]_{num}]$
N --> [estudiant_]
VP --> aux verbal
aux --> $[[\text{past}]_t]$
verbal --> $[\text{aprobar}]_v$det N
det --> $[[[1\_]_{art}]_{def}]$
N --> [curso COMP352]

After the tranformation rules have been applied the sentence obtained is,

treinta estudiantes aprobaron el curso

Then, the procedure READ-SPEAK will rewrite the sentence as
(1) (2) trein-ta es-tu-dian-tes (1) // a-pro (2) ba-ron el cur-so COMP352(1)

Notation: __ stressed syllable, (.) intonation level, // pause marker

b) There are 8 sections of COMP352, Do you want a list of them?
The DBMS and the sentence generator together generate P2 as a response to Q2.

S --> NP VP
NP --> [1_] [curso COMP352]
VP --> $[[\text{nonpast}]_t]$ verbal
verbal --> $[[\text{tener}]_v[[8]_{ord}]$ N
N --> [seccion_]
S --> Q [tu] aux verbal
aux --> $[[\text{nonpast}]_t]$
verbal --> [querer] NP
NP --> [un_] [lista]
comp --> [de] [el_]

The output from the transformational grammar is

el curso COMP352 tiene 8 secciones. quieres una lista de ellos?

This response will be rewritten by the procedure READ-SPEAK as
(1) el (2) cur-so COMP352 tie-ne 8 sec-cio-nes (1) // (2) quie-res u-na lis-ta (1) // de (3) e-llos

For the generation of P1 and P2 we have used a transformational grammar for Spanish given in [32,pp.228-243]. The selectors that control the application of the production rules for generating a sentence will derive the necessary information from the user query(Q), DBMS(S), response to the query obtained by searching the database(R), knowledge source and the data dictionary(K), and the grammar(G).

V.-SUMMARY:

We have presented brief introductions to the different concatenative units used for speech synthesis and to the Orthography and Phonology of Spanish. Syllabication of the written words in Spanish, placement of stress and pause markers, and intonation level markers have been discussed. The objective of placing these markers is to produce a natural sounding speech that could be useful in database applications for machine to man communication.

REFERENCES

[1] Dudley, H., and Tarnoczy, T.H., "The speaking machine of Wolfgang von Kempelen" J. Acoust. Soc. Am., vol.22, pp.151-166, 1950.
[2] H.Wakita and D.J.Broad, "Models and methods for speech synthesis, in The Nervous System, D.B.Tower, ed., vol.3, Human Communication and Its Disorders, New York:Raven, pp.459-467, 1975.
[3] J.L.Flanagan, Speech Analysis Synthesis and Perception, New York:Springer, 2nd ed., 1972.
[4] L.R.Rabiner and R.W.Schafer, "Digital Techniques for computer voice response: implementations and applications", Proc. IEEE, vol.64, no.4, pp.416-432, 1976.
[5] J.Allen, "Synthesis of speech from unrestricted text", Proc. IEEE, vol.64, no.4, pp.433-442, 1976.

[6] I.G.Mattingly, "Syllable synthesis", Haskins Laboratories : Status Report on Speech Research SR-49, 1977.
[7] O.Fujimura and J.Lovins, "Syllables as concatenative units", in [9] pp.107-120.
[8] M.J.Hunt, M.Lennig, P.Mermelstein, "Experiments in syllable-based recognition of continuous speech", IEEE Int. Conf. Acoust. Speech and Signal Proc., pp.880-883, 1980
[9] A.Bell and J.Hooper, Syllables and Segments, Amsterdam:North-Holland, 1978.
[10] P.M.Bertinetto, C.Miotti, S.Sandri, and E.Vivalda, "An interactive synthesis system for the detection of Italian prosodic rules. CSELT Rapporti tecnici, vol.v, no.5, pp325-331, Dec. 1977.
[11] R.T.Gagnon, "Votrax real time hardware for phoneme synthesis of speech", IEEE Int. Conf. Acoust. Speech and Signal Proc., pp.175-178, 1978.
[12] Elovitz H.S., Johnson R.W., McHugh A., and Shore J.E., "Automatic translation of English text to phonetics by means of letter-to-sound rules", Naval Research Laboratory, Washington D.C. Report No.7948, 1976.
[13] McIlroy, M.D., "Synthetic English speech by rule", Bell Telephone Laboratories, March 1974.
[14] G.Kielczewski, "Digital synthesis of speech and its prosodic features by means of a microphonemic method", Speech Communication with Computers, L.Bolc, ed. London:Macmillan pp.183-206, 1978.
[15] Mezzalama,M. and Rusconi,E. "A general system for synthesizing speech", Speech Communication, G.Fant, ed., pp.307-314, 1974.
[16] B.A.Sherwood, "Fast text-to-speech algorithms for Esperanto, Spanish, Italian, Russian, and English", Int. J. Man-Machine Studies, vol.10, pp.669-692, 1978.
[17] M.Chafcouloff, Vingt Cinq Annes de Recherches en Synthese de la Parole, Paris:CNRS, 1976.
[18] A.Juilland and E.Chang-Rodriguez, Frequency Dictionary of Spanish Words, The Hague:Mouton, 1974.
[19] R.Seco, Manual de Gramatica Espanola, Madrid:Aguilar, 1976.
[20] T.Navarro, Manual de Pronunciacion Espanola, Madrid:Rayscar, 1963.
[21] T.Navarro, Studies in Spanish Phonology, University of Miami Press, 1968.
[22] P.Delattre, Comparing the Phonetic Features of English, French, German and Spanish, Verlag, 1965.
[23] B.Malmberg, Phonetics, New York:Dover, 1963.
[24] F.Berdichevsky, G.Murillo, C.Cuttler, "A microcomputer based speech synthesizer which speaks Spanish", IEEE Int. Conf. Acoust. Speech and Signal Proc., pp.917-920, 1979.
[25] T.Winograd, Understanding Natural Language, Edinburgh:University Press, 1972.
[26] J.Mylopoulos, A.Borgida, P.Cohen, N.Roussopoulos, J.Tsotsos and H.Wong, "TORUS - a natural language understanding system for data management", 4th Int.J.Conf. Artificial Intelligence, Vol,1, pp.414-421, Tbilisi USSR, Sept 1975.
[27] D.Waltz, "Natural language access to a large data base: an engineering approach", 4th Int.J.Conf. Artificial Intelligence, Vol.2, pp.868-872, Tbilisi USSR, Sept 1975.
[28] F.Fallside and S.J.Young, "Speech output from a computer-controlled water-supply network", Proc. IEE, Vol.125, No.2, pp.157-161, Feb. 1978.
[29] D.R.Hill, "Spoken language generation and understanding by machine: a problems and applications oriented overview", NATO ASI 1979.
[30] H.C.Lefkovits, Data Dictionary System, Q.E.D. Information Sciences 1977.
[31] J.E.Grimes, Network Grammars, University of Oklahoma, 1975.
[32] R.L.Hadlich, A Transformational Grammar of Spanish, New Jersey:Prentice-Hall, 1971.