Generation and Synthesis of Broadcast Messages

L.F. Lamel, J.L. Gauvain, B. Prouts, C. Bouhier, R. Boesch

LIMSI-CNRS
BP 133
91403 Orsay cedex, FRANCE
{lamel,gauvain}@limsi.fr

VECSYS
le Chêne Rond
91570 Bievres, FRANCE

ABSTRACT

In this paper we present a system designed by VECSYS, in collaboration with LIMSI, to generate and synthesize high quality broadcast messages. The system is currently undergoing evaluation tests to automatically generate and broadcast messages about weather and airport conditions. The generated messages are alternately broadcast in French and in English.

This paper focuses on the synthesis aspects of the system which synthesizes messages by concatenation of speech units stored in the form of a dictionary. The texts corresponding to the messages to be synthesized and broadcast are automatically generated by a server, and transmitted to the synthesis component. Should the system be unable to generate the message from existing entries in the dictionary, the system permits a human operator to record the message which is subsequently broadcast.

INTRODUCTION

Air traffic control services are responsible for assuring air travel under the safest and most regular conditions possible. As such, they are charged with regulating the flow of air traffic, of avoiding accidents, and of transmitting timely information and alerts during landing and takeoff (airport control) and for flights within the control area. It is the air traffic controller, located in a regional control center who assures direct contact with the airplane pilots. One important aspect of the communication is the transfer of information about meteorological conditions, local and regional, and information particular to the airport such as runways open for takeoff and landing, and conditions on the runways.

The system DIVA, (DIffusion Vocale Automatique des informations meteorologiques), has been developed for the diffusion of vocal information to pilots about weather conditions (VOLMET) and about airport conditions (ATIS). The VOLMET messages are automatically composed and synthesized by a central server (Aviation Civile d'Orleans-Bricy) for several regions in France and transmitted by dedicated phone lines to the regional control centers. The regional control centers broadcast these regional meteorological conditions to aircraft in flight within the control region. The ATIS messages provide information about local weather conditions and conditions particular to the airport. They are destined to aircraft taking off or landing, and are composed from local weather reports and from observations from the air traffic controllers. Currently these messages are recorded as needed by a controller on magnetic tape, and are continually broadcast. The messages are updated as needed throughout the day by different operators, with different voice qualities which appears to the pilots as inconsistencies in the qualities of the recordings.

The goal of this work was to improve the quality of the synthesis for VOLMET and to automatically generate messages for ATIS. The system has been designed by the VECSYS Company, under a contract with the STNA (Service Technique de la Navigation Aérienne). A prototype system with permanent diffusion of ATIS messages accessible by telephone was delivered in January 1993. The system to automatically generate and broadcast messages about weather and airport conditions should be in functional use by the end of 1993.

This paper focuses on the synthesis aspects of the project, which were developed in collaboration with LIMSI. While many techniques for speech synthesis have been investigated over the years (see [3, 1, 5, 8] for a review), in this work since the aim is high quality message generation within a limited vocabulary domain where all of the vocabulary elements are defined in advance, synthesis by waveform concatenation has been chosen so as to provide the most natural vocalised message. The system synthesizes messages by concatenation of speech units which are stored in the form of a dictionary. The speech units can be words, subword units, or short phrases that are automatically extracted from previously recorded task-dependent sentences, covering the application vocabulary and syntax. The texts of the information messages to be synthesized and broadcast are automatically generated by a server, and transmitted to the synthesis component via an X25 connection. In the cases where a message cannot be generated from the existing entries in the dictionary, the system permits a human operator to record the message which is subsequently broadcast.

Synthesis is performed by concatenation of the dictionary units according to the text string provided by the message generator. Each text entry in the dictionary may have several associated prerecorded signals, so as to be able to account for contextual variations. The sequence of speech units are selected from the stored entries, using dynamic programming to optimize the overall quality of the synthesized message, by taking into account the phonetic and/or word context, the pitch of successive units, and punctuation markers. Messages may be broadcast in French, in English, or alternating French and English.

APPLICATION SPECIFICATION

The information messages to be broadcast are of the type VOLMET or ATIS. The VOLMET message provide pilots with weather conditions in the control region, and the ATIS messages provide information specific to the conditions at the airport, destined to pilots of aircraft taking off or landing at the airport. The messages are continually broadcast on designated frequencies appropriate for the aircraft, and can also be heard over the telephone. The synthesized messages are limited to a maximum duration of 16 minutes. The length of the messages and the language (French, English, or alternating French and English) are parameters of the configuration which can be specified for each message individually.

The composition of the information to be broadcast is obtained from numerical codes received from a central server and may be augmented by spoken information by an on-site operator in conditions where the necessary dictionary units do not exist. For each application (weather conditions or airport conditions) the dictionary contains about 400 entries, and the number of possible sentences that can be generated is on the order of 10^{12}. An example message from VOLMET is given in Figure 1 and an example message for ATIS is given in Figure 2.

MESSAGE GENERATION UTILITY

For the VOLMET information, the messages are automatically composed by a central server, and numeric codes are transmitted via an X25 connection to the synthesis component. The diffusion of the synthesized message is automatically controlled. In the case of ATIS, two modes of operation are supported. The first is automatic message composition, where the operator has the possibility to verify the message, and to add a complementary message if needed. The second mode is manual, where the operator records the message to be diffused, and initiates the broadcast after verification.

BIARRITZ , NEUF-HEURES . TROIS UNITE ZERO DEGRES , ZERO SIX NOEUDS . SUPERIEURE-A-DIX-MILLE-METRES . TROIS-OCTAS CUMULUS DEUX TROIS ZERO ZERO-PIEDS . QUATRE-OCTAS STRATOCUMULUS CINQ TROIS ZERO ZERO-PIEDS . CINQ-OCTAS ALTOCUMULUS HUIT TROIS ZERO ZERO-PIEDS . TEMPERATURE UNITE CINQ , POINT-DE-ROSEE UNITE ZERO . QNH-UNITE-ZERO UNITE HUIT . NOSIG

BIARRITZ , ZERO-NINE . THREE ONE ZERO DEGREES , ZERO SIX KNOTS . MORE-THAN-TEN-THOUSAND-METERS . THREE-OCTAS CUMULUS TWO THREE ZERO ZERO-FEET . FOUR-OCTAS STRATOCUMULUS FIVE THREE ZERO ZERO-FEET . FIVE-OCTAS ALTOCUMULUS EIGHT THREE ZERO ZERO-FEET . TEMPERATURE ONE FIVE , DEW-POINT ONE ZERO . QUEBEC-NOVEMBER-HOTEL-ONE-ZERO ONE EIGHT . NOSIG

Figure 1: Example message for VOLMET application in French and in English.

The following commands are available to the operator for message generation:

- selection of the channel for diffusion
- selection of language of the message
- synthesis of a message from a specified text (numerical form) or recording of a spoken message by the operator
- verification of the recorded message with simultaneous presentation of the text material
- authorization to diffuse the resident message
- authorization to destroy the current message, stopping its diffusion

For both applications, the duration of the message and the language can be specified. The current message is repeatedly broadcast until a new message is received. The new message will start to be transmitted after the end of a current message cycle. Since it is important that the information that is transmitted be accurate, in cases of uncertainty, a message is stopped, and a default message, identifying the station is transmitted until a more accurate message is ready for transmission.

ICI MERIGNAC, INFORMATION UNIFORME ENREGISTREE A UNITE HEURE UTC.

PISTE AU DECOLLAGE 2 7.

SUR PISTE 2 7 FLAQUES D'EAU SUR CENT POUR CENT DE LA PISTE 2 CENTIMETRES.

FREINAGE DOUTEUX.

NIVEAU DE TRANSITION CINQ ZERO.

ATTENTION POUR LES DEPARTS DES SPECIFICATIONS SERONT DONNEES SUR LES FREQUENCES DE CONTROLE.

SI PERTE DE CONTACT SUR 113 DECIMALE 15, PASSER SUR UNITE 0 8 DECIMALE 9 5.

ATTENTION PRESENCE D'OISEAUX SUR LE TERRAIN.

VENT 0 8 0 DEGRES, 9 NOEUDS.

PORTEE VISUELLE DE PISTE INFERIEURE A 4 0 0 METRES.

PLUIE.

NEBULOSITE 6 OCTAS CUMULONIMBUS INFERIEUR A 4 0 0 METRES.

TEMPERATURE UNITE 2 DEGRES.

POINT DE ROSEE 6 DEGRES.

QUEBEC FOXTROT ECHO 1 0 1 7.

INFORMEZ LA ROCHELLE, DES LE PREMIER CONTACT QUE VOUS AVEZ RECU L'INFORMATION UNIFORME.

THIS IS MERIGNAC, INFORMATION UNIFORM RECORDED AT 1 UTC.

TAKE OFF RUNWAY 2 7. ON RUNWAY 2 7 WATER PATCHES ON ONE HUNDRED PERCENT OF THE RUNWAY 2 CENTIMETERS.

BRAKING ACTION UNRELIABLE.

TRANSITION LEVEL FIVE ZERO.

CAUTION FOR DEPARTURES CLIMB INSTRUCTIONS WILL BE ISSUED ON CONTROL FREQUENCIES.

IF RADIO FAILURE ON 113 DECIMAL 15 CHANGE TO 1 0 8 DECIMAL 9 5.

CAUTION BIRDS ON THE AIRFIELD.

WIND 0 8 0 DEGREES, 9 KNOTS.

RVR LOWER THAN 4 0 0 METERS.

RAIN.

NEBULOSITY 6 OCTAS CUMULONIMBUS LOWER THAN 4 0 0 METERS.

TEMPERATURE 1 2 DEGREES.

DEW POINT 6 DEGREES.

QUEBEC FOXTROT ECHO 1 0 1 7.

INFORM LA ROCHELLE, ON INITIAL CONTACT THAT YOU HAVE RECEIVED INFORMATION UNIFORM.

Figure 2: Example message for ATIS application in French and in English.

DICTIONARY OF SPEECH UNITS

The dictionary of speech units for each application (VOLMET or ATIS) contains entries for all elements needed to generate all possible vocal messages. The entries consist of variable-size units selected so as to minimize the amount of material to be stored while assuring high quality speech generation. Thus, whenever possible, phrases are concatenated to provide continuity when the contents are not variable, and smaller units are used to compose the variable portions. There are separate dictionaries for French and for English. For each application there are about 400 entries, which allow the construction of the respective messages. Since there is overlap in the vocabularies for the two tasks, some of the dictionary elements are able to be shared by the two applications.

The dictionary elements contain the International Aeronautical Alphabet for the letters (alpha, bravo, charlie, delta, ...,), and for the numbers. There are some differences in phraseology for English and French. For example, in English all numbers are spoken as digits strings: 1055 is spoken as "1"(one) "0"(zero) "5"(five) "5"(five), whereas in French the same string is pronouced "10"(dix) "55"(cinquante-cinq). The other elements can be subclassed into carrier phrases, airport names, identifiers, terms describing the weather and runway conditions, and times.

The dictionary entries were obtained by recording sample phrases, sentences, and messages, and extracting the units to be used for concatenation. The sample phrases were generated from a grammar that specifies the syntax for each application. The syntax was created using a utility, REBUS[4], which provides a graphical interface for the user to build a context-free grammar. In generation, the utility ensures that each vocabulary item appears in the list of sentences, and that all possible types of sentences are represented. The program also attempts to minimize the number of sentences necessary to cover the syntax and vocabulary. In order to account for the variability observed in pronunciation due to local stress and sentential position, several occurrences of each dictionary entry were included in different positions.

The list of sample phrases were then read aloud by a bilingual speaker, who read the sentence lists for French and English. The recorded signal was automatically segmented into the dictionary entries using a speech recognizer[6] to align the text prompt with the signal. The segmentations are used to define the start and end of the recorded speech associated with each dictionary entry. Associated with each dictionary entry are orthographic and phonemic transcriptions which are used in selecting the units used to generate a message. Initially minimal phrases were recorded so as to cover the application vocabulary. However it was found that, particularly for French, the naturalness of the synthesized messages was greatly improved when examples of longer messages were recorded.

SYNTHESIS BASED ON CONCATENATION

The synthesis of a given transmitted message entails division of the message into dictionary entries, selection of the best instance of the speech units for the sequence of entries, and concatenation of the signal corresponding to the selected speech units with local smoothing to avoid discontinuities. Synthesis by concatenation[2] of pre-stored speech units has served as the basis of several systems already developed by VECSYS in collaboration with LIMSI. The technique has been particularly successful in the context where the synthesis serves to verify phrases automatically recognized. One particular application in daily use at MSI is for the resynthesis of recognized numbers for control of inventory. The user reads numbers such as the manufacturing identification numbers, the quantity in stock and the sale price, which are recognized, and resynthesized for verification. For this application the dictionary entries consist of disyllables,[1] so as to handle the contextual variation observed across word boundaries and to avoid the unnaturalness obtained when silences are inserted between words. Using these units, 121 disyllable models can be used synthesize any sequence of digits, and 137 models can generate the numbers from 0 to 9999 in French.

In order to select the "best" possible sequence of dictionary units to produce a given sentence, a sequence cost is estimated for all the possible sequences corresponding to the sentence. This cost is obtained by summing local costs including pitch discontinuity costs, intra-segment pitch variation costs, phone context costs, word context costs and end of sentence pitch cost. The sequence of segments having the overall smallest cost is chosen to synthetize the sentence. The search of this "best" sequence is efficiently performed by using a dynamic programming algorithm. The weights associated to each cost type can either be fixed by hand (based on subjective quality measures) or can automatically be obtained from a list of training sentences. In the later case, the weights are obtained by minimizing the sum of all the training sentence costs.

Silence segments of variable length are used to insure a good rhythm thus improving the naturalness of the message. To avoid discontinuities due to concatenation, segment boundaries are placed at a beginning of a pitch period for voiced segments and continuity is assured for the signal and its first derivative. Windowing the segment boundaries is often necessary to achieve this goal.

DISCUSSION

An early version of the synthesizer did not use dynamic programming to select the optimal sequence of speech units and also did not make use of pitch information. It was quickly observed that even relatively small differences in pitch between successive units were very disturbing if this difference was in opposition to expected, natural pitch variations such as end of phrase declination. The use of a few very simple rules (implemented as dynamic programming costs) to avoid inappropriate rises and falls of the pitch greatly aided the naturalness of the concatenated message.

Early assessment of synthesis quality indicated the need for recording entire example messages, rather than simply short phrases covering the vocabulary items. This greatly improved the naturalness of the synthesis, particularly for French, by providing the correct intonation. It was also found that judicial use of variable length silence aids the naturalness of the perceived message.

The complete system has been delivered and is being integrated with the national server. Subjective evaluation of the synthesis quality is underway.

REFERENCES

[1] Calliope (1989), *La Parole et Son Traitement Automatique*, Paris: Masson.

[2] S.E. Estes, H.R. Kerby, H.D. Maxey, R.M. Walker (1964), "Speech Synthesis from Stored Data," *IBM J. Res. Develop.*, **8**, 2-12. (reprinted in *Speech Syntheses*, J.L. Flanagan, L.R. Rabiner, eds., Pennsylvania: Dowden, Hutchinson & Ross, Inc. 1973)

[3] J.L. Flanagan, L.R. Rabiner, eds. (1973), *Speech Syntheses*, Pennsylvania: Dowden, Hutchinson & Ross, Inc.

[4] J.L. Gauvain (1990), "Le système de reconnaisance *AMADEUS:* Principe et algorithmes," *LIMSI internal report*, June 1990.

[5] D.H. Klatt (1987), "Review of text-to-speech conversion for English," *J. Acoust. Soc. Am.*, **82**(7), 737-793.

[6] L.F. Lamel, J.L. Gauvain (1992), "Experiments on Speaker-Independent Phone Recognition Using BREF," *Proc. IEEE ICASSP-92*, San Francisco, CA, **S1**, 557-560.

[7] H. Singer, J.L. Gauvain (1988) "Connected speech recognition using dissyllable segmentation," *Fall meeting of the Acoust. Soc. of Japan.*

[8] C. Sorin (1991), "Synthèse de la Parole à Partir du Texte: Etat des Recherches & des Appliations," *Actes de Deuxiemes Journées Nationals de GRECO PRC Communication Homme-Machine*, Toulouse.

[1] A disyllable is defined from the midpoint of one vowel to the midpoint of the next vowel.[7]