

Automatic Generation of Hypotheses for Automatic Diagnosis of Pronunciation Errors

Salah Eldeen Hamid^{1,2,3} , Mohsen Rashwan^{2,3}

¹Department of Electrical Engineering, Higher Technological Institute
(10th of Ramadan City, Egypt)

²Department of Electronics and Communication Engineering, Cairo University.
Giza, Egypt.

³The Engineering Company for the Development of Computer Systems; RDI.
171 Al-Haram main st., 6th floor, Giza, Egypt
{salah, mrashwan}@rdi-eg.com

Abstract

This paper describes the use of a rule based system for generation of pronunciation variants as a component of a speech-enabled computer aided pronunciation learning (CAPL) system. This CAPL system is a part of a computer aided recitation of the holy Qur'an training system. It generates the most probable pronunciation error hypotheses that are fed to a hidden Markov model (HMM)-based speech recognizer in order to test them against the spoken utterance. It also generates mapping information to determine the appropriate location for the feedback of each candidate hypothesis.

1. Introduction

CAPL has received a considerable attention in recent years. Many research efforts have been done for improvement of such systems especially in the field of second language teaching (Herron et al 1999, Franco et al 1999; Cucchiaroni et al 1998; Witt 1999). In this system we target automatic training for correct recitation of the holy Qur'an for Arabic speakers. Shortage of experienced teachers in most of environments and/or lack of sufficient time at learner's side makes this system a highly demanded system. It not only helps students to learn how to recite the holy Qur'an but also helps them to correct their errors in formal Arabic pronunciation.

Improving the performance of CAPL system requires detailed and accurate knowledge about causes, locations and types of mispronunciations in the user's utterance. This knowledge also helps greatly to increase accuracy of automatic speech classifier by limiting tested hypotheses only to probable ones. If compared for example, to the results of an unconstrained phone-loop recognizer, a considerable reduction in general recognition uncertainty can be achieved (Bonaventura et al 2000).

Although knowledge about probable mispronunciations is highly domain dependent as it almost completely depends on both target application and the status of targeted users. The framework for identifying types and effects of probable mispronunciations and procedure for generation of search lattices and determination of error locations is still common for different applications and targeted users.

This framework includes classification of pronunciation errors and realization of a matching rule based system for generation of pronunciation hypotheses. The rules can be ranked according to error relevance or impact, thus providing additional cues for the selection between competing hypotheses (Bonaventura et al 2000). It also enables the system to retract the best hypothesis and finds the most probable cause of the error and thus giving the most helpful feedback to the learner.

In this paper we will use the SAMPA Arabic phonetic alphabet (SAMPA) to represent transcription of pronunciation variants.

2. Types of Recitation Errors

Recitation errors can be classified according to two major aspects (Namely: 1- Cause of the error which determines the suitable feedback and 2- Effect of the error which determines method of detection).

2.1 Classification of Recitation Errors Based on Cause of Error

Knowing the cause of an error mainly determines the suitable and subjective feedback to the learner, which guides the learner to the cause of his error and may give him the missing information and/or practice needed to overcome it.

2.1.1 Errors Caused by Colloquial Pronunciation:-

In these errors the equivalent formal Arabic pronunciation doesn't exist in mother colloquial dialect.

Examples of such errors:-

- Egyptian colloquial Arabic doesn't contain the phonemes T (ت) and D (د), s (س) and z (ز) is pronounced instead.

For these types mentioning the difference between the correct pronunciation and the mistaken one is the appropriate feedback for most of these error types.

2.1.2 Errors Caused by Inaccurate Articulation

These errors usually happen due to insufficient practice and/or knowledge of correct articulation of phonemes. It

usually happens because of adjacency of phonemes with distant place of articulation that results in difficulty in correct pronunciation of phoneme with weaker characteristics. (Non-emphatic letter mistakenly emphasized if adjacent to emphatic letter).

Examples for these errors:-

- (الْتَار) (?anna:r') becomes (?anna':r') (where 'a' is emphatic a)
- (تَصِير) (tas'i:r) becomes (t'a's'i:r)

Correction of these types of errors is much harder than first type because:-

1. It needs practice to get acquainted to the correct pronunciation (which is difficult).
2. It is very difficult for learners to diagnose this type of mispronunciations by themselves.
3. It changes the style of speaking of some words than Arabic colloquial dialect.
4. This type of mispronunciations happens very frequently.

The appropriate feedback for this type of errors must include accurate diagnosis of the error, its cause, illustrative examples of pairs of the correct and the mistaken pronunciations, also a link to a practice session on this subject may be very helpful.

2.1.3 Errors Caused by Ignorance of Recitation Rules

Because the holy Qur'an has some recitation rules that differ than formal Arabic pronunciation, even a well trained formal Arabic speaker shall make recitation errors.

Examples of these errors are:-

- Pronouncing non-vowelled noon (Manifestation) when it should be concealed: (مَنْ زَكَّهَا) (man:zakka:ha:) becomes (manzakka:ha:)
- Pronouncing a vowel followed by glottal stop without extra lengthening as it is compulsory lengthening as in (السَّمَاء) (?assama::?) becomes (?assama:?) (Where a:: stands for a double length a:).

Correction of this type of errors needs full explanation (with illustrative examples) of the rule and its application at the current verse. If the correct pronunciation contains a phoneme that doesn't exist in formal Arabic it should be described in details with the help of pronunciation examples.

2.1.4 Errors Due to Over-Generalization of Recitation Rules

Because general rules of recitation has exceptions in some locations in the holy Qur'an (usually none or few for each rule), even experienced speakers may be ignorant of all of these exceptions.

Correction of this type should inform the learner of the existence of the exception and a link to the information about other locations where that exception exists.

2.2 Classification Based on Effect of Recitation Error

Good understanding of the acoustic differences between a correctly-pronounced recitation and a wrongly-pronounced one is the key for selection of efficient mispronunciation detection method.

2.2.1 Errors Affecting Phonetic Transcription

That includes phone insertions, deletions and substitutions, such as

- Insertion of glottal stop "همزة وصل" in (عن النبأ) (?'aninnaba?) so it becomes (?'an?annaba?)
- Deletion of phoneme r "ر" at end of utterance as in (خبير) (xabi:r) so it becomes (xabi:)
- Substitution of phoneme T "ث" by phoneme s "س" seen as in (ثواب) (Tawa:b) so it becomes (sawa:b)

HMMs well-trained on the acoustic features of each phoneme will be a sufficient detection tool for these error types.

2.2.2 Errors Affecting Phoneme Duration

That includes lengthening of short phonemes or vise versa such as:

- Lengthening vowel (i) in (مالك يوم الدين) (ma:likijawmiddi:n) so it becomes (ma:liki:yawmiddi:n)
- Shortening of compulsory lengthening in (السماء) (?assama::?) so it becomes (?assama:?)

Ordinary acoustic HMM is not a sensitive discrimination tool between models with similar acoustic features but differs only in duration. So HMMs will be deployed only in phonetic segmentation of the utterance then speaking-rate normalization and a duration classification algorithms will be used. The detailed description of these modules is beyond the scope of this paper.

3. Generation of Pronunciation Errors Hypotheses

The generation process of pronunciation error hypotheses should be flexible enough to deal with error hypothesis addition, deletion and overlapping of probable mispronunciations.

3.1 Module Architecture

Figure (1) shows a block diagram for the search lattice generation module.

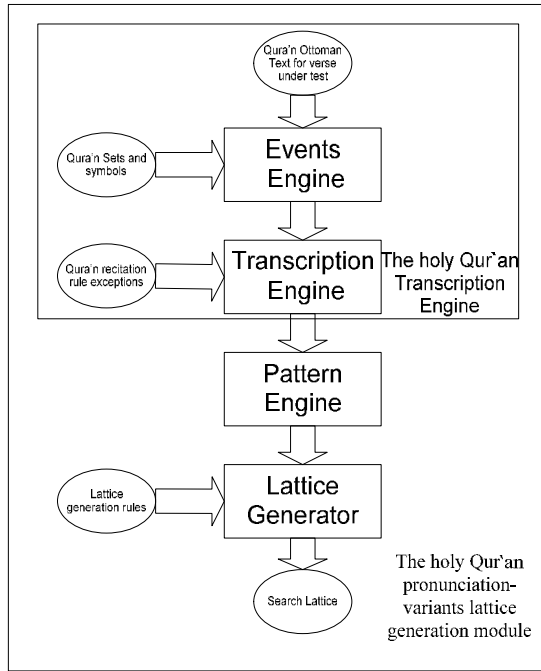


Fig (1) Block diagram for the holy Qur'an pronunciation variants lattice generator

The lattice generator is built on basis of RDI's holy Qur'an transcription engine. The transcription engine is built in the form of multi-layer event driven modules. This architecture was selected to enable any higher level analysis module to use this core engine and benefit from its results. That what was actually done with our pronunciation variants lattice generator.

The events engine scans the input holy Qur'an Ottoman text searching for symbols and features and at each probably pronounced character it generates its code, its pronunciation status and its acoustic characteristics (such as voicing, place of articulation, nasalization and aspiration).

The transcription engine analyzes those codes and characteristics and generates the corresponding correct phonetic transcription according to the holy Qur'an recitation rules and their exceptions.

The pattern engine gathers all the information from proceeding layers and generates pronunciation patterns at probable pronunciation locations. These pronunciation patterns are used for matching with pronunciation variants rules at the lattice generator.

The lattice generator sorts the matched rules descending with their error relevance or impact then all rules resulting in the same phoneme sequence are omitted except the first one. Finally the lattice is generated with remaining pronunciation variants in a format suitable to the speech recognizer. Also a mapping file is generated that holds the locations of the suitable feedbacks for each pronunciation variant.

3.2 Lattice Unit

Because we need our system to generate user helpful feedbacks and that most Qur'an learners are not familiar with phonemes we choose our lattice unit similar to the one used in traditional methods of the holy Qur'an recitation teaching "EL-KOTTAAB".

In those methods the text is divided to basic units

1. Consonant + short vowel (CV) like: ba bi bu sa si so ... etc.
2. Consonant + long vowel (CVV) like: ba: bi: bu: sa: si: su: ... etc.
3. Non-vowelled consonant (C) like: b s d r l .. etc.
4. Repeated consonant +short vowel (CCV) like: bba bbi bbu ssa ssi ssu ...etc.
5. Repeated consonant + long vowel (CCVV) like: bba: bbi: bbu: ssa: ssi: ssu: ... etc.
6. Non-vowelled repeated consonant (CC) like: qq RR nn ... etc.

Applying these units to the transcription of:-

(بِسْمِ اللّٰهِ الرَّحْمٰنِ الرَّحِیْمِ) (bismilla:hirra`X\ma:nirra`X|i:m)

It becomes:-

(bi)(s)(mi)(lla:)(hi)(rra`)(X\)(ma:)(ni)(rra`)(X|i)(m)

3.3 Lattice Generation Rules:-

Studying the types of recitation errors (Safaqusi 1974) we found that they usually depend mainly on current unit, the nearest proceeding unit and the nearest succeeding unit. Effect of far units is limited to the error rank and can be ignored safely on the assumption that all pronunciation hypotheses are generated.

3.3.1 Rule format

The rules take the form:-

IF

For Current unit

Phoneme = Ph_c, Vowel type = V_c, Vowel length = L_c, Doubled = S_c, Pronounced = P_c and Concealment = E_c

AND

For previous unit

Phoneme = Ph_p, Vowel type = V_p, Vowel length = L_p, Doubled = S_p and Pronounced = P_p

AND

For next unit

Phoneme = Ph_N, Vowel type = V_N, Vowel length = L_N, Doubled = S_N and Pronounced = P_N

THEN

Add a path to recitation error lattice with the following parameters

Error Code = C, Error Type = T, Error word =W, Error frequency rate = F, Phoneme = Ph, Vowel type=V, Vowel length = L, Doubled = S, Pronounced = P and Concealment = E

Where

- S, P and E are binary flags.

- E is used to indicate a lattice branch that continues over the succeeding unit.
- C holds error code in the error table for further information and feedback.
- T is the type of the error (Recitation, Vowel type or Length error).
- W is a modification word used to produce non standard pronunciations.
- F holds the frequency of the error (common, less common, fair, rare)

3.3.2 Rule Databases

Benefiting from the efforts of (Safaqusi 1974) in documenting most probable pronunciation errors made by speakers while reciting the holy Qur'an, we managed to construct a database for rules of pronunciation errors in the holy Qur'an recitation. We also added our efforts to add any noticeable recitation error. The database size reached 663 matching rules.

This is the database of recitation errors that we have recognized; the architecture of the system enables us to freely use a subset of this database to produce limited scale realizations of the system. A probable application is to generate "Vowel Type" errors only for a realization dedicated to beginners.

This database is connected via error code to two other databases. The first database is the feedback database which holds the coloring codes, the readable feedbacks and the audible feedbacks.

The second database also holds links between each specific recitation error and relevant holy Qur'an recitation rule(s) (Hosary 2002). This database is used to filter the pronunciation errors to concentrate on specific recitation rules for a given lesson.

4. Results

In order to reach an evaluation for the complete CAPL system, an evaluation database was needed that contains a set of voice utterances representing the recitations of randomly selected users of the system of different gender, age and proficiency combinations. These utterances were evaluated by a number of language experts –in separate sessions-, and labeled with the actual pronounced phonemes. For ambiguous speech segments experts were allowed to write all acceptable judgments in their opinions. After each expert has finished, all experts' transcriptions are summed to produce a list of all the judgments accepted by the experts. Afterwards, a final group session is held where all experts discuss each error and they can agree on either to keep all the judgments or choose one or more of them that is to correct any transcription errors that may be generated by them.

This database is used to evaluate the system, by comparing the system responses with human experts' transcriptions.

So the database will consist of a set of utterances, and all acceptable transcriptions for each utterance. It means that some pronounced segments may have more than one acceptable judgment. So each judgment has four possibilities:

- 1- Correct (accepted by all human experts).
- 2- Identified pronunciation error (all human experts reported the same error).
- 3- Not Perfect (human experts disagreed whether to reject or accept the pronunciation). That can happen when pronunciation is not perfectly correct.
- 4- Wrong with unidentified error type (human experts agreed that a pronunciation error exists but disagreed on its type). That can happen when the user makes complex or undocumented errors.

The system's HMM-based speech recognizer associates each decision it makes with a corresponding confidence score that is used to choose the suitable feedback response to the learner. When the system suspects the presence of a pronunciation error with low confidence score the system has some alternate responses:-

- 1- Omit the reporting of the error at all (which is good for novice users; because reporting false alarms discourages them to continue learning correct pronunciation).
- 2- Ask the user to repeat the utterance because it was not pronounced clearly.
- 3- Report the existence of an unidentified error and ask the user to repeat the utterance (which is better for more advanced users than ignoring an existent error or reporting wrong type of pronunciation error).
- 4- Report most probable pronunciation error (which –if wrong- can be very annoying to many users).

At each suspected error, deciding which feedback to adapt is dependant on the error's confidence score and the system thresholds setting. System thresholds are set according to progress level of the current user.

Table (1) shows a sample application of the evaluation system. The table shows the distribution of occurrences of judgment-response pairs. These values were calculated using the system settings for novice users.

As we see in the table (1), for correct speech segments the system yielded "Repeat Request" for about 9.7% of the total correct segments. That is because they had low confidence score below the computed threshold, and the system gave a repeat request to avoid the possibility of false alarms.

For Wrong speech segments which constitute 8.2% of the data, the system correctly identified the error in 52% of the errors, reported unidentified errors for 4% and repeat requests for 24% of the errors. The system made false acceptance of 17% of the total errors.

		Human judgement				
System Judgement		Correct	Wrong	Not Perfect	Wrong with Unidentified Error Types	Total
	Correct	80.89%	1.43%	1.07%	0.00%	83.39%
	Wrong with same error type	0.00%	4.29%	0.18%	0.00%	4.46%
	Wrong With Wrong Error Type	0.00%	0.18%	0.00%	0.18%	0.36%
	Repeat Request	8.75%	1.96%	0.71%	0.00%	11.43%
	Wrong with Undefined Error	0.00%	0.36%	0.00%	0.00%	0.36%
	Total	89.64%	8.21%	1.96%	0.18%	100.00%

Table 1 : Occurences Distribution

5. Conclusion

A module for the automatic generation of pronunciation hypotheses for pronunciation errors automatic diagnosis was built as a component in a complete computer-aided holy Qur'an recitation learning system.

Recitation errors were classified to reach the suitable feedback and detection methods. Then a recitation error matching rules database was built that holds the most probable recitation errors.

An important advantage of the developed system, that it enables safe removal/addition to the rule database. This advantage enables system developers to tailor a version of the system for a specific type of errors, or to add new pronunciation variants if needed (for example for non native Arabic speakers).

6. Acknowledgments

Special thanks are posed to The Engineering Company for the Development of Computer Systems (RDI) <http://www.rdi-eg.com> for its support of the pioneer application of CAPL technology in holy Qur'an recitation learning. We gratefully acknowledge their support for this research.

We must mention valuable efforts of speech technology and linguistic support teams. Special thanks to Naim Abdelghani for his sincere efforts in authoring system feedback and to Ahmed Ragheb for putting all his domain knowledge under our request. We would also wish to express many thanks to Yassir Hifny and Ossama Abdel-Hameed for implementing the transcription engine.

References in Arabic

- (Hosary 2002) الحصري، محمود خليل، أحكام قراءة القرآن الكريم، مكتبة السنة، ط1: 1423 هـ، 2002 م.

- (Safaquisi 1974) الصفاقسي، أبو الحسن علي بن محمد النوري، تنبيه الغافلين وإرشاد الجاهلين عما يقع لهم من الخطأ حال تلاوتهم لكتاب الله المبين، المطبعة الرسمية للجمهورية التونسية، 1974 م.

References in English

- Bonaventura P., Herron D., Menzel W. (2000) Phonetic rules for diagnosis of pronunciation errors. Proceedings of Konvens 2000 (Conference on Natural Language Processing), pp. 225-230, Ilmenau, Germany, 9-12 October 2000.
- Cucchiari, C. & Strik, H. & Boves, L. (1998), Automatic pronunciation grading for Dutch. Proceedings STiLL '98, Marholmen, Sweden, pp.95-98 .
- Franco H, Neumeyer L, Ramos M, and Bratt H, (1999) Automatic Detection of Phone-Level Mispronunciation for Language Learning, Proc. of Eurospeech 99, Vol. 2, 851-854, Budapest, Hungary.
- Herron, D., Menzel W., Atwell E., Bisiani R., Daneluzzi F., Morton R., Schmidt J. A. (1999) Automatic localization and diagnosis of pronunciation errors for second-language learners of English, Proc. 6 th European Conference on Speech Communication and Technology, Eurospeech 99, Budapest, pp. 855-858.
- SAMPA "computer readable phonetic alphabet" <http://www.phon.ucl.ac.uk/home/sampa/home.htm>,
- Witt, S. (1999) Use of Speech Recognition in Computer-Assisted Language Learning. PhD thesis, Cambridge University Engineering Department, Cambridge, UK.

This document was created with Win2PDF available at <http://www.daneprairie.com>.
The unregistered version of Win2PDF is for evaluation or non-commercial use only.