(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
20 December 2012 (20.12.2012)

WIPO | PCT

(10) International Publication Number
WO 2012/173582 A1

(51) International Patent Classification:
G09B 5/06 (2006.01) *G09B 19/06* (2006.01)
G10L 13/00 (2006.01)

(21) International Application Number:
PCT/TR2012/000092

(22) International Filing Date:
6 June 2012 (06.06.2012)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
2011/05998 17 June 2011 (17.06.2011) TR

(71) Applicant *(for all designated States except US)*: **SESTEK SES VE ILETISIM BILGISAYAR TEKNOLOJILERI SANAYII VE TICARET ANONIM SIRKETI** [TR/TR]; ITÜ Ayazağa Kampüsü Koru Yolu Arı-2, Teknokent Binası A Blok No:A4-4, Maslak 34469 Sariyer Istanbul (TR).

(72) Inventor; and
(75) Inventor/Applicant *(for US only)*: **BOYRAZ, Bayram** [TR/TR]; ITÜ Ayazağa Kampüsü Koru Yolu Arı-2, Teknokent Binası A Blok No:A4-4, Maslak 34469 Sariyer Istanbul (TR).

(74) Agent: **DESTEK PATENT, INC.**; Konak Mah. Lefkose Cad. NM Ofis Park B, Blok No: 36 / 5 BEŞEVLER NİLÜFER, 16110 Bursa (TR).

(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) Designated States *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:

— *with international search report (Art. 21(3))*

— *before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments (Rule 48.2(h))*

(54) **Title**: USING SPEECH SYNTHESIS FOR LANGUAGE TRAINING WITH PICTURE SYNCHRONIZATION

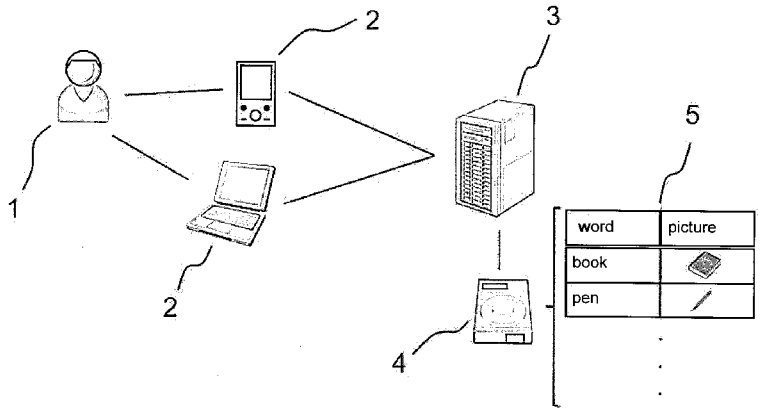

Fig. 1

(57) **Abstract**: The invention relates to the method for using speech synthesis in language training with picture synchronization, including the process step of providing feedback to the client (2) about what word to be synthesized at that moment in a continuous text flow, by means of the speech synthesis motor; characterized by comprising the following consecutive steps; the text to be synthesized is sent to the server (3) by the client (2), the server (3) looks for the most suitable picture for each word from picture database (5) in the memory unit (4), the server (3) sends the most suitable pictures found in picture database (5) to the client (2), the client (2) shows the pictures delivered by the server (3) to it synchronically on its screen, together with the pronunciation of each word respectively.

WO 2012/173582 A1

DESCRIPTION

USING SPEECH SYNTHESIS FOR LANGUAGE TRAINING WITH PICTURE SYNCHRONIZATION

Technical Field

The invention relates to the use of speech synthesis (Text-To-Speech) for language training, by showing a picture expressing the word to be synthesized during synthesis.

The invention, particularly, relates to the use of speech synthesis with picture synchronization for language training, characterized by the following consecutive steps; the text (word, sentence, paragraph, or the whole textbook) to be synthesized is sent to the server by the client, the server looks for the most suitable picture for each word in the text from picture database in the memory unit, the server sends the most suitable picture found in picture database to the client, the client shows the picture delivered by the server to it synchronically on its screen, together with the pronunciation of the each word in sequence.

Prior Art

The use of speech synthesis for language training is not a new method. Most people use speech synthesis for learning how to pronounce the words in a foreign language. Since most software highlights the unit of text during synthesis, the user uses such software so as to memorize/learn the pronunciation of the words. However, the user does not learn anything about the meaning of the text synthesized.

In avatar applications used in speech synthesis (WO2008096099), it is aimed to give a facial expression, depending on the characteristic of the phoneme to be

synthesized. Such applications relates to the fact that the avatar takes a sentimental expression according to the structure of the word to be synthesized during speech synthesis. The only thing illustrated on the screen is a "talking head".

Within the state of the art, the applications performing speech synthesis are not intended for language learning. For example, in avatar applications, while the word "pen" is being synthesized, not a pen image, but a talking head is shown on the screen.

Also, the current applications do not offer pronunciation, spelling and the related picture of the word all at once along with a continuous text flow. For example TR200505402 is based on learning the words located in the database by showing the meaning and the related picture of the word on different parts of the screen, while pre-recorded audio files are played from the speakers. The invention aims language training however, the main difference is, the system does not include speech synthesis, and it can not be used to pronounce any word sequence. Therefore the system can only play the audio for the words present in the database and can not support any given continuous text. KR20090132692 also represents the content of a text as pictures, however the system does not include any speech synthesis. Consequently, it does not offer any voice data as pronunciation of the words.

JP2008129434 offers speech synthesis however the system only reflects atmosphere corresponding to human sensibility such as bright or dark. It has no interest in presenting pictures relevant to text.

As a consequence, while speech synthesis is only used for pronunciation in language learning, the present invention is the only method that aims to allow the users to truly understand and speak a foreign language by means of speech synthesis.

Object of the Invention

The invention, inspired by the current state, aims to solve the above mentioned disadvantages.

The object of the invention is to use the speech synthesis for language training with picture synchronization.

A further object of the invention is to show the pictures expressing each word in a continuous text flow separately, not just the phoneme.

A further object of the invention is to facilitate language training/learning by way of the voice-picture match in the memory of the user, thanks to the fact that the user hears the pronunciation of the synthesized word, as well as seeing the picture illustrating that word on the screen.

A further object of the invention is to provide the user to understand what is spoken by making the pictures used in speech synthesis appear in the mind of the user, hearing the same word used in speech synthesis, simultaneously.

A yet further object of the invention is to allow the user to speak a foreign language as well as understanding it, thanks to the voice-picture matches recorded in the mind of the user during speech synthesis.

A further object of the invention is to facilitate the user to speak the foreign language, by making the voice corresponding to the objects seen during daily life/the concepts encountered/the images made up appear in the mind of the user.

To achieve the above mentioned purposes, the invention is a system and method for using speech synthesis in language training, with picture synchronization, including a speech synthesis motor providing feedback to the

client about which word to be synthesized instantaneously in a continuous text flow, and being composed of a client synchronically showing the picture delivered to it by the server on the screen along with the pronunciation of the word, a memory unit where picture database is provided, and a server which finds the most suitable picture for the word to be synthesized from the database in the memory unit and sends this picture to the client; wherein the system is characterized by the following steps which are performed by said elements: words to be synthesized is sent to the server by the client respectively, the server looks for the most suitable picture for each word from picture database in the memory unit, the server sends the most suitable pictures found in picture database to the client, the client shows the pictures delivered by the server to it synchronically on its screen, together with the voice pronunciation of each word respectively.

The structural and characteristic aspects and all the advantages of the present invention will be more clearly understood by means of the following figures, and the detailed description written with references to these figures, and therefore, while making an evaluation, these figures and the detailed description should be taken into account.

The Figures Helping the Invention to Be Understood

Fig. 1 is a schematic diagram illustrating the components used in the preferred embodiment of the system according to the invention wherein speech synthesis is used in language training with picture synchronization.

Description of Part References

1. User
2. Client
3. Server
4. Memory unit

5. Picture database

The drawings do not necessarily need to be scaled, and the details that are not required to understand the present invention may be disregarded. Apart from that, at least substantially identical elements and the elements having at least substantially identical functions are shown with the same reference numeral.

Detailed Description of the Invention

In the detailed description, the preferred embodiments and process steps of the system according to the invention wherein speech synthesis is used in language training with picture synchronization will be explained only for the subject to be better understood.

In the system according to the invention where speech synthesis is used in language training with picture synchronization, at least one client (2) having at least one speech synthesis motor thereon, at least one server (3), and at least one memory unit (4) where picture database (5) is provided are used (Fig. 1). Speech synthesis motor is positioned on the client (2) and has the same function as that in the prior art speech synthesis systems. The role of the speech synthesis motor is to provide feedback to the client (2) about what word to be synthesized instantaneously in a continuous text flow. Within the system, preferably, mobile phones, pda (pocket computers), tablet computers, or computers are used as client (2) and server (3); and the memory unit (4) is preferably hard disk connected with the server (3) (Fig.1). In the method where speech synthesis is used in language training with picture synchronization, first, the text to be synthesized in the client (2) is provided, or the user (1) enters said text or word by himself/herself. During speech synthesis, when the user (1) wants to synthesize a text, i.e. when he/she wants to get the pronunciation and what is meant by the words in the text, she/he states through the client (2) that she/he wants to perform said process. Feedback is provided to the client (2) about what word to be synthesized instantaneously in a continuous text flow, by

means of the speech synthesis motor positioned on the client (2). The client (2) sends words in the text to be synthesized to the server (3) separately. The server (3) looks for and finds the most suitable picture for each word to be synthesized in picture database (5) in the memory unit (4) connected thereto. The server (3) sends the found picture to the client (2) again. The client, in turn, shows the delivered pictures on the screen for the user (1), in synchronization with the pronunciation of each word.

The client (2) may be the same machine as the server (3), while it may be different machine (mobile phone, pda, tablet computer, computer) from the server (3). It is aimed by this flexibility that any user with or without space problems can use the system.

In the system for using speech synthesis in language training with picture synchronization, since the pictures will be shown during the synthesis of the texts, the system aging is prevented. That is, the user (1) will not be shown the same picture list constantly and will not have to listen to a constant announcement. Instead, with the ever-changing world, the user (1) will listen to changing updated texts, and suitable pictures for those texts will be shown. In time, the only thing to be renewed/improved will be picture database (5).

The system for using speech synthesis in language training with picture synchronization is a system based on showing the picture illustrating the synthesized text. For example, if the synthesized word is "pen", a pen image will be shown on the screen of the client (2); similarly, if it is "bag", a bag image will be shown.

Moreover, the primary object is language training in the system and method for using speech synthesis in language training with picture synchronization. In the system, a picture concerning the meaning of the synthesized word is shown. The picture to be shown on the screen will be the picture of an object if the word to be synthesized refers to an object, will be a picture expressing a verb if the

word is a verb, and will be a picture expressing a sentiment if it is sentiment. For example, while the word "pen" is being synthesized, the picture to be shown on the screen will be independent of the context where the word "pen" is used. The same pen image will be displayed on the screen, whether the word "pen" is used in a text expressing happiness, or in a text expressing sadness.

The system and method for using speech synthesis in language training with picture synchronization directly deal with the meaning of each word appearing in text.

The system and method for using speech synthesis in language training with picture synchronization according to the invention operates similarly, for any user who wants to learn English as a foreign language. For instance, while the word "book" (meaning "kitap" in Turkish) is being synthesized, the server (3) will find the book image best expressing the meaning of the "book" in picture database (5), and send it to the client (2). The client (2), in turn, will show the book image to the user (1) in synchronization with its pronunciation. Thus, while the word "book" is being displayed on the screen, the user will not only hear the pronunciation of the word "book" from the client (2), but she/he will also see the picture best expressing the word "book", i.e. the book image, on the screen of the client (2), in synchronization with the pronunciation of said word. Thus, the voice-picture matches recorded in the mind of the user (1) during speech synthesis will allow the user to speak a foreign language, as well as understanding it. The voice corresponding to the objects the user sees in his/her daily life, the concepts she/he encounters, and the images she/he makes up will appear in the mind of the user (1); hence, the user will be able to speak the foreign language.

CLAIMS

1. The invention is a system for using speech synthesis in language training with picture synchronization, including at least one speech synthesis motor that provides feedback to the client (2) about what word to be synthesized at that moment; characterized in that it comprises

   - at least one client (2) displaying the pictures delivered by the server (3) to the user (1), in synchronization with the pronunciation of each word in a continuous text flow,
   - at least one memory unit (4) where picture database (5) is provided,
   - at least one server (3) that finds the most suitable picture for each word to be synthesized from picture database (5) in the memory unit (4) and sends that picture to the client (2).

2. The system for using speech synthesis in language training with picture synchronization according to Claim 1; characterized in that

   - said client (2) is a mobile phone, a pda (pocket computer), a tablet computer, or a computer.

3. The system for using speech synthesis in language training with picture synchronization according to Claims 1 and 2; characterized in that

   - said server (3) is a mobile phone, a pda (pocket computer), a tablet computer, or a computer.

4. The system for using speech synthesis in language training with picture synchronization according to Claims 1 to 3; characterized in that

- said memory unit (4) is a hard disk.

5. The invention is a system for using speech synthesis in language training with picture synchronization, including the process step of providing feedback to the client (2) about what word to be synthesized in a continuous text flow, by means of a speech synthesis motor; characterized in that it comprises the following process steps in respective order:

- the text to be synthesized is sent to the server (3) by the client (2),
- the server (3) looks for the most suitable picture for each word in the text, from picture database (5) in the memory unit (4),
- the server (3) sends the most suitable pictures found in picture database (5) to the client (2),
- the client (2) displays the pictures delivered to it by the server (3), on its (2) screen in synchronization with the pronunciation of each word.

6. The method for using speech synthesis in language training with picture synchronization according to Claim 5; characterized in that

- a mobile phone, a pda (pocket computer), a tablet computer, or a computer may be used as said client (2).

7. The method for using speech synthesis in language training with picture synchronization according to Claims 5 and 6; characterized in that

- a mobile phone, a pda (pocket computer), a tablet computer, or a computer may be used as said server (3).

8. The system for using speech synthesis in language training with picture synchronization according to Claims 5 to 7; characterized in that

- a hard disk is used as said memory unit (4).

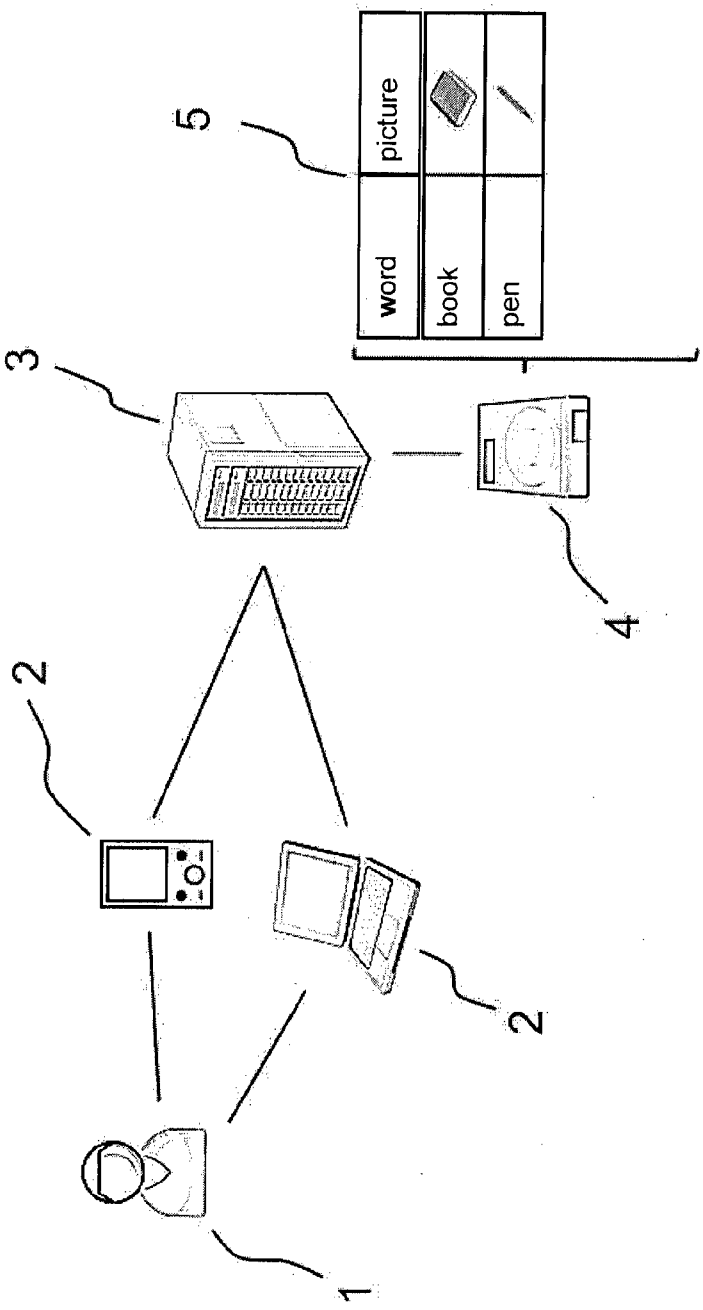

Fig. 1

INTERNATIONAL SEARCH REPORT

International application No
PCT/TR2012/000092

A. CLASSIFICATION OF SUBJECT MATTER
INV. G09B5/06 G10L13/00
ADD. G09B19/06

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
G09B G10L

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
EPO-Internal, WPI Data, COMPENDEX, INSPEC, IBM-TDB

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	US 4 695 975 A (BEDRIJ OREST J [US]) 22 September 1987 (1987-09-22) column 2, line 9 - line 13 column 2, line 47 - line 48 column 3, line 39 - line 44 column 4, line 31 - column 5, line 49 column 8, line 3 - line 6 figures 1, 12-16	1,4,5,8

X	WO 01/97198 A1 (TAN BAHN JEE LOLA [SG]) 20 December 2001 (2001-12-20) page 1, line 3 - page 2, line 17 page 5, line 21 - line 25 page 8, line 5 - line 16 claims 1-4	1,4,5,8

	-/--	

[X] Further documents are listed in the continuation of Box C.

[X] See patent family annex.

* Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
5 November 2012	16/11/2012

Name and mailing address of the ISA/	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Fax: (+31-70) 340-3016	Geißler, Christian

INTERNATIONAL SEARCH REPORT

International application No
PCT/TR2012/000092

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	US 5 885 083 A (FERRELL JOE G [US]) 23 March 1999 (1999-03-23) column 4, line 44 - line 57 column 7, line 34 - line 65 figure 4	1,4,5,8

X	US 2006/188852 A1 (GORDON GAYLE E [US] ET AL) 24 August 2006 (2006-08-24) paragraphs [0051], [0052], [0058], [0059], [0062], [0068], [0074] figures 14A-14D	1-8

X	US 2011/107217 A1 (SCHWARZ MARGERY KRAVITZ [US]) 5 May 2011 (2011-05-05) paragraphs [0008], [0013], [0062], [0067] figure 1	1-8

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No
PCT/TR2012/000092

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
US 4695975 A	22-09-1987	CA 1239705 A1	26-07-1988
		DE 3581894 D1	04-04-1991
		EP 0182460 A1	28-05-1986
		JP 61101183 A	20-05-1986
		US 4695975 A	22-09-1987
WO 0197198 A1	20-12-2001	AU 5437900 A	24-12-2001
		CN 1454376 A	05-11-2003
		HK 1062069 A1	20-04-2007
		MY 135740 A	30-06-2008
		WO 0197198 A1	20-12-2001
US 5885083 A	23-03-1999	CA 2202105 A1	09-10-1997
		EP 0801370 A1	15-10-1997
		IL 120622 A	17-02-2000
		JP 3249764 B2	21-01-2002
		JP 10039741 A	13-02-1998
		TW 411422 B	11-11-2000
		US 5885083 A	23-03-1999
US 2006188852 A1	24-08-2006	US 2006188852 A1	24-08-2006
		WO 2006071574 A2	06-07-2006
US 2011107217 A1	05-05-2011	NONE	