

⑦

208-213

WorldUnion 多国语言实时自动交流系统

刘 炜, 孙雅如, 何晓阳

(西北大学 计算机科学系, 陕西 西安 710069)

TN912.3

TP391.42

摘 要: WorldUnion 多国语言实时自动交流是为解决世界范围内语言交流障碍提供的一种新的可行性方案, 它把现有成熟的语音识别、语言翻译、语音合成技术用软件集成的方法进行二次开发, 其功能是接收、识别一国语音, 实时自动地把其翻译成另一国语言并用相同的音频合成朗读。此系统在网络通信、信息交流等领域都可应用, 市场前景广阔。

关键词: 语音识别; 语言翻译; 语音合成; 软件集成; 动态数据交换

中图分类号: TN918.64; O322 **文献标识码:** A **文章编号:** 1000-274 X (1999)03-0208-06

World Union

语言交流系统

在高度信息化的社会中, 语言是世界信息交流中最大的障碍。人们或是掌握多种自然语言, 其学习成本巨大, 或是邀请翻译, 其费用高昂, 且可靠性、保密性不高。直接利用语言翻译软件也要经过文字手工输入、书面语词句切分、视觉阅读屏幕等复杂过程, 不能适应现代社会高效率要求。

针对以上问题, WorldUnion 多国语言实时自动交流系统(以下简称交流系统)对现有成熟软件进行重用和二次开发, 其功能是在接收一国语音的同时自动地将其翻译成另一国语言, 并用相同音频合成朗读, 使语言不通的双方能实时地自由交流。此系统适用于国际间科学技术交流、信息通讯、商务合作等有语言障碍的场合, 其社会效益和经济效益都是不可估量的。

1 WorldUnion 多国语言实时自动交流系统各组成部分的原理

汉语和英语是使用人口最多的自然语言, 以下以汉英相互交流为例, 介绍交流系统各组成部分的原理。

1.1 语音识别

语音识别时预先将标准的单词元素对应的隶属度函数存入标准语音库。用户语音信号通过带通滤

波器和低通滤波器组进行声波频谱分解, 提取单位时间内声波所含的频谱参量、零交叉率和线预测参量等语音特征参量, 将这些参量作为音素抽出。经时间的非线性扭曲处理, 再根据特征参量与标准语音进行反复多次校核和差距测量, 求出差距最小的单词作为识别结果^[1]。

表1 计算机与人的翻译功能比较

Tab. 1 Comparison of Function of Computer and Human in Translation

翻 译 家	计 算 机
1 汉语语音声波通过听觉器官传至大脑	语音识别
2 在大脑中把汉语转换成英语	语言翻译
3 指挥发音器官发出相应英语语音	语音合成

要提高语音识别的正确性, 就要使计算机具有模糊模式识别能力。二值化时频图型模糊匹配法是模糊逻辑应用于语音识别的一个成功尝试, 它将 TSP 图(time spectrum pattern)进行二值化处理, 每一行峰值为1, 其余取0, 这样, 既保留并突出了发声单词的共振频率, 又大大缩减了数据量和识别时间^[2]。

1960年, Denes 等人研究成功了第一个计算机语音识别系统。目前, 以大词汇量、非特定人连续语音识别为标准, 一些语音识别系统已达到商品化阶段。IBM 公司投入巨资, 历时26年, 推出中、英、法、德、日等8个国家语言的识别产品, 例如 Viaoice 4.0

收稿日期: 1998-05-10

作者简介: 刘 炜(1977-), 男, 上海人, 西北大学本科生。

中文语音识别系统。其特点是非特定人连续语音识别,具有强大的学习功能,通过用户发声来训练调整隶属度函数,在使用中不断提高识别率,平均识别率可达 90%以上,配上专用的输入设备,每分钟可输入 150 个汉字。其衍生产品瑞驰安 Dutty++ 中文语音识别系统还提供 Dutty++ API 接口,方便用户进行二次开发。

1.2 语言翻译

机器自动翻译是把一种语言的文本文件自动翻译成另一种语言的文本文件。它首先需要两个语言的互译词典作为翻译系统的重要知识来源,所有被翻译的元素信息,包括单词的各种语义、搭配以及相关知识,都应包含在这个海量词典中。其次,因为语言翻译中的个性规则远多于通用规划,所以多义词和词的搭配都要以规则的形式收入词典。另外,近期发展起来的语料库学为全文翻译的研究打开了一个突破口,其基本思想是构建一个经过加工的语料库,将输入的句子和语料库中的句子比较,从而得到相似度最大的句子作为译文输出。语料库在应用中不断完善、扩展,精确度也不断提高。若互译的两种语言同属一个语系(如英语和法语),由于语言规则相似,翻译精确度就会高得多。

目前全文翻译软件以东方快车 2000、中软译星、通译科技翻译等为代表。天津大通开发的通译第五代产品就采用了上述“基于规则,辅以语料”的新型研究方法,以 2 000 多条规则为基础建立一套完善的规则处理机制。同时,它亦设立了庞大的语料库和 24 个专业词库以储存大量的相关信息,英汉互译,译文可读率达到 85%以上。

1.3 语音合成

语音合成是将一个文本文件根据不同的语音合成原理转换为数字语音信息,再经过 D/A 转换为模拟语音信号,送扬声器发声。目前通行 3 种合成系统:第一种是参数合成系统,提取相应的语音参数加以存储,合成时将语音参数译码,再经 D/A 转换、滤波、放大合成语音;第二种是基频叠加系统,由于是调用说话人的录音材料来拼成语句,这样单字和两字组的自然度就高,成本也比较经济;第三种是按规则合成系统,根据声位和声调变化规则,以最小的合成单元合成出无限的自然语音。例如,利用正在发展的神经计算技术来合成语音。它由人工神经元互相联接成网,以充分并行的方式进行信息处理,在一个学习过程中借助于改变神经元联接的权值而得到正确的发音^[3]。

现在已商品化的语音合成软件有很多,例如 LED 汉字语音卡仿真软件,它可将一篇中文文稿用标准的普通话朗读出来,用户可以对语速、音量、声调(男女声)等进行选择和控制;Smooth 英语自动朗读软件可发出标准英语语音,短小精悍(只有 233KB 容量)。以上语音合成软件均不需要声卡,操作使用极其方便,朗读语音清晰悦耳。

2 WorldUnion 多国语言实时自动交流系统的构成

以上介绍了语音识别、语言翻译、语音合成的原理及相关技术与成果。从软件开发成本和软件维护两方面进行考虑,有效利用软件重用技术必将提高软件的生产率。WorldUnion 就是基于当前成熟的技术成果进行的二次开发,而不是耗资耗时的重复开发。它把语音识别、语言翻译、语音合成 3 种技术以相辅相成、互相联系方式有机地结合成一体,构成一套语言实时自动交流系统。

2.1 硬件及软件要求

因为语音识别要求配置较高,所以最低配置为多能奔腾 166 MHz 处理器,32M 内存,16 位声卡,大容量外存贮器以及话筒、扬声器和相应接口。为了用户携带方便,可选用笔记本或手持式电脑。

以上所介绍的商品化的语音识别、语言翻译、语音合成系统中的软件内核都可选用。若要形成多国语言交流,只需加入这些国家语言的语音识别、语音合成以及它们相互之间的翻译软件即可,在此不再赘述。由于要进行软件集成,操作系统最好选用微软公司的 Windows。

2.2 模块化设计思想

目前,模块化方法已为所有工程领域所接受。模块化设计带来了许多好处:一方面,模块化设计降低了系统的复杂性;另一方面,它推动了系统各个部分的并行开发。所以在交流系统中,把语音识别、语言翻译和语音合成作为系统的 3 个组成模块。

模块化设计的最终目标,是建立起独立性强的内聚低耦合的模块。由于各识别、翻译、合成系统本来都是无联系相对独立的软件产品,所以这些模块内核的功能是单一明确的,都属于内聚性高的功能内聚模块。然而,在交流系统中这些模块间的连接方式却是简单的,属于数据耦合。由于限制了只通过参数传递数据,模块间耦合性很低,设计、测试以及维护其中任何一个模块,都不会对系统中其他模块

产生影响^[4]。

在交流系统中,模块间的通信数据放在缓冲区中,进一步降低了模块间的耦合度。缓冲区将减少模块间互相等待的时间,把缓冲区作为数据通信的媒介,那么一个模块执行的速度、频率等问题一般不影响其他模块的设计。

2.3 系统处理流程和实现技术

本系统处理流程为:一国语音通过话筒和带通滤波器组被频谱分解成数字信号,此信号通过语音识别模块加工变换为该语言的文本信息,此文本再经翻译模块加工成另一国语言的文本信息,最后把译后的文本送交语音合成模块合成发音。

那么,现在的问题是语音识别、语言翻译及语音合成 3 大模块在有限的内存中,如何同时存放和怎样实时并发执行? 经过加工后中间数据通过什么方式传输和交换?

应用最广泛的 Windows 操作系统支持内存保护模式和程序重装入技术,在保护模式中,所有段都是可移动的。当需要时,Windows 可以在内存中移动可移动段,以给另一个程序提供足够的空间。这样,Windows 可以在合成 3 模块时同时为它们分配内存并按需装入一部分代码。数据从一个模块的缓冲区传输至另一模块的缓冲区时,若处理这些数据的代码不在内存时,Windows 用程序重装入技术就可以动态地将缺少的代码用 LRU 算法换入内存。

当交流系统的 3 大模块同存于内存后,就可以开始数据的加工和传输了。DDE 是一种动态数据交换的消息协议,它使用 Windows 消息和共享的内存,使相互合作的应用程序能够自动进行数据交换。因发送和接收程序在交换数据时都必须遵守 DDE 协议,所以若想把 3 大模块集成一体,每一模块都必须加以改进,利用 DDEML API 接口来给这些应用系统添加支持 DDE 的功能。

有些 DDE 消息中含动态分配数据块句柄,所有给对方应用程序的动态内存均须用 Globallock 函数来分配,并在函数中给出 GMEM-DDESHARE 参数。当发送数据的应用程序释放 CPU 时,Windows 在接收应用程序自己的全局内存页中分配一个新块,然后 Windows 将原来内存中内容拷入新近分配到的内存块。接收数据的应用程序获得 CPU 后,执行对数据的操作。

在设计交流系统主模块时可使用支持 DDE 协议的 Windows 编程语言(如 VC++, Delphi 等)。在变换型系统结构中把主模块作为 DDE 客户,由它负责人机界面设计,以及作为媒体传递中间结果;把语音识别、语言翻译、语音合成子模块作为 DDE 服务器,负责进行数据加工。通过主模块和 3 个子模块间动态联接和数据交换实现软件集成。系统在内存中的工作流程框图如下(以中英交流为例)。

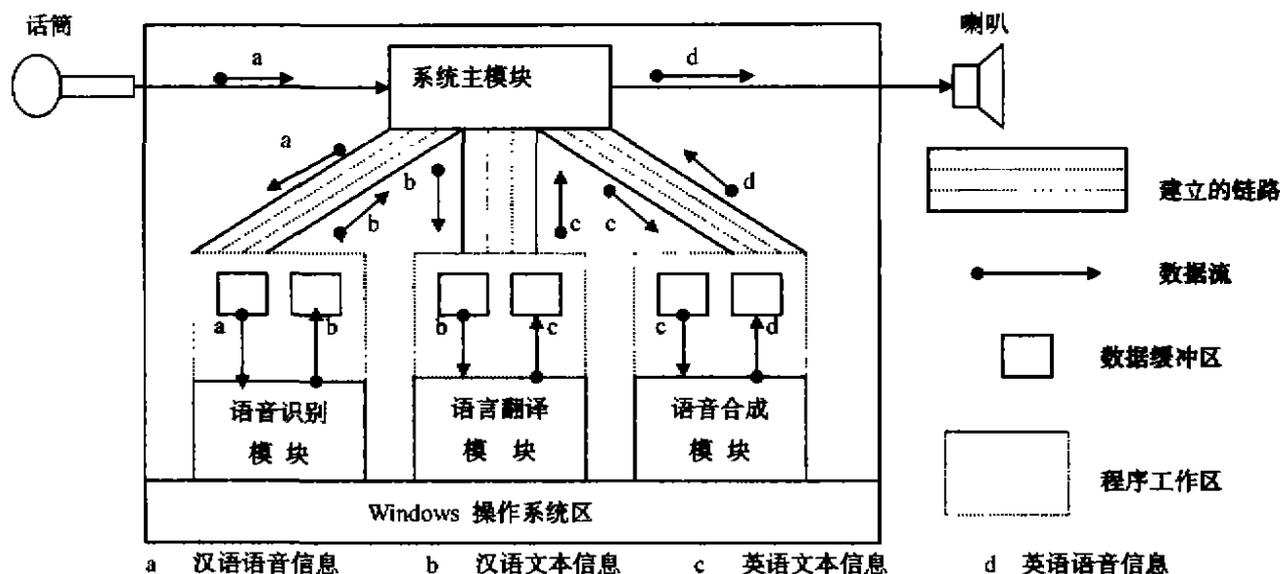


图 1 WorldUnion 的实现过程

Fig. 1 Flow Diagram of WorldUnion

在 DDE 动态数据交换时主要用到的 DDE 消息功能如下:

用 WM-DDE-INITIATE 建立主模块与子模块之间的链路;

用 WM-DDE-REQUEST 实现主模块客户向子模块服务器的缓冲区要求加工后的中间结果;

用 WM-DDE-POKE 实现主模块将数据中间结果拷贝至下一服务器的缓冲区中;

用 WM-DDE-EXECUTE 向下一服务器发送命令让其加工缓冲区中的数据;

用 WM-DDE-TERMINATE 结束一次动态数据交换^[5]。

本系统还可使用流水线技术,使识别语音、翻译语言以及合成语音 3 级子任务能在流水线的各个阶段并发地执行。从理论上讲,一个 3 级过程段的流水线处理 n 个长度相似的句子需要的周期数为

$$T_3 = 3 + (n - 1) = n + 2。$$

流水处理的加速比为 $C_3 = \frac{n \times 3}{n + 2} \approx 3 (n \gg 3)$ 。

这也就是说,使用 3 级线性流水线几乎可以将其速度提高 3 倍,从而使整个系统的实时性大为改善。

由于 WorldUnion 是让不同国家的用户使用,所以应设计一个多语通用的人机界面。系统功能应具备选择语言种类(在多种语言中选择两种)、交流方向(单向或双向)等。在双向交流时,例如汉英交流完后,按功能键应自动把汉语语音识别转换为英语语音识别、翻译换为英译汉、英语语音合成转换为汉语语音合成,即可实现英汉交流。系统强大的学习功能使其在用户使用过程中不断提高语音识别率和翻译忠实度,越用越精确。

3 WorldUnion 多国语言实时自动交流系统各组成间的联系

即使 3 大模块的精确度都达到 95%,组合在一起若不经调整,精度就下降到 85%左右。所以,交流系统把 3 模块集成为一体时要删除一些重复的部分(如各自的界面)和系统不需要的功能(如语音合成软件中的文本编辑功能),抽取这些应用软件的内核,建立起它们之间的联系,使它们相辅相成,互相促进。这样做不仅是系统集成中提高精确度的途径,而且是提高整个系统自然度和仿真度的关键。

3.1 语音识别词的切分、翻译和语音合成间的联系

汉语书面语是按句连写的,它滤掉了人们思想中的关键信息——词间间隙。在翻译句子和合成语音时都要用到词的切分知识(把按句连写转换为按

词连写)。词的正确切分是书面翻译正确与否的决定因素,它的任何错误都可能使译文产生严重的偏差和错误。分词正确也可使音节的变调处理确定下来,使合成语音自然度提高^[6]。

只有用韵律知识才能解决歧义句的问题。口语中,词间有“顿挫”(按词说出)。显然,人口语中的间隙可以忠实地表达说话人的思想。WorldUnion 可以用语音输入中的间隙信息来较圆满地解决词的切分问题。例如人们在说“羽毛球拍卖完了”时,若意思是“拍卖羽毛球”,语音就会在“球”和“拍”之间有短暂的间隙,传感器捕捉到声音的强度在“球”和“拍”之间较长时间变弱,通过计算就可以把“羽毛球”和“拍卖”两词切分开来;若意思是“卖羽毛球拍”则说话时会在“拍”和“卖”之间有间隙,同理会很准确地把“羽毛球拍”和“卖”切分开来。总之,根据语音识别分词的结果来提高翻译的忠实度和确定多音字的语音及变调处理是可行的。

在识别和翻译结合时还要注意口语和书面语的区别,比如有时口语省略主语,而在形成书面语时则必须予以保留。这类问题都是在系统集成时需要解决的^[7]。

3.2 翻译结果和语音合成之间的联系

在今天的词典软件市场上,英汉互译和在线式男女即时发音的“会说话的双向词典”已成为一种商业标准。但是,词典终究是词典,它只能用手输入且只能读出单个词的语音。若把单词语音直接拼接而不作过渡协同处理就会产生切断语音、无语音音变、无升降调、听感不连贯等问题。

为翻译文本合成语音时,要使合成的语音像口语一样连贯自然,词语中的起头和结尾音节,就要有过渡性质的韵律音变,它主要表现在声调的升势和降势,音长的连续和停顿及语气的轻重。要根据自然语言中的协同发音规律把各音节间的变化及音高、音长和音强三项韵律特征处理好。自然度处理可以使得合成语音的清晰度与可懂度达到更高程度^[8]。

3.3 语音识别和语音合成之间的联系

人们都希望自己说的话经过交流系统处理后不要“变味”。如果一男子的语音输入后合成出一女子的声音则会让人啼笑皆非。所以,在语音识别阶段时,就要同时存入该语音的频率、音强和语速等相关特征参数,以此来调整语音合成时的合成参数,使合成的语音与说话人的语音相符。这不仅让用户能听到如同自己说出的外语语音,也能让听众辨别出是谁在使用系统。以上这些改进均力求使交流系统的

自然度和仿真度达到让用户满意的程度。

4 WouldUnion 多国语言实时自动交流系统可行性研究

4.1 技术可行性

第一,现有的语音识别、语言翻译、语音合成技术已达到相当高的精确度,而大量的人力和物力仍在投入更新、扩充这些技术成果,所以相关技术发展是可以支持这个系统的;第二,现在的硬件水平已能够使本系统初步达到实时、自动的目的,而硬件的性能和处理速度仍在高速增长;第三,IBM 公司在开发语音识别产品时,并没有把它做成一个独立的应用程序,而是把它做成一个“语音内核”,以动态链接库的方式来运行,使之易于集成。

4.2 经济可行性

第一,由于利用已有的先进技术成果进行集成,不需重复研究,开发成本低,设计周期短;第二,相关技术被不断并行地更新换代,只需移植相应更先进的技术成果即可系统升级,系统维护成本低;第三,硬件价格正迅速降低,有利于系统普及;第四,交流系统的使用范围无国界,不受本软件开发语言限制,在世界范围内任意两国都可应用它交流,所以此产品可推广国内外,有形效益和无形效益都是巨大的,潜在市场前景乐观。

5 展 望

WorldUnion 多国语言实时自动交流系统的研究,具有明确的应用目的,这套系统只需再加入少量的软、硬件,就可以发挥更大的用途。

(1)一语识别多语输出:由一国语言语音识别,联合多套语言翻译和语音合成系统而组成。如在国际会议环境下,一中国人演讲,中文语音识别后把识别结果分别送往中英、中法、中德、中日等语言翻译和语音合成系统,不同国籍的人戴上耳机可实时接听语音翻译。这类一转多系统还可安装在国际广播、通讯领域,把本国信息传送到世界各地。

(2)多语识别一语输出:这类系统可以利用在本

国信息接收系统上,不同国家的信息被接收后,通过多语识别,翻译,再合成该国语音,可直接向听众播放。

(3)改善残疾人的交流环境:有手疾的朋友可用语音识别代替文字输入,并且可以把自己“说”成的文章自动翻译成各种语言的文章;失聪的朋友使用交流系统时只需加入一液晶显示屏,别人说的话经翻译(或不翻译)在屏幕上显示;口哑的朋友可以使用联机手写识别系统,用笔写代替键入,使他们能听到自己“说”出的本国语或外语;失明的朋友可以用扫描仪扫描书本报纸,再用语音合成朗读;使用盲文的朋友,还可使用本系统结合盲文机器码和盲文点字的输入输出与外国朋友自由交流;等等^[9]。

(4)WorldUnion 在网络通信上的应用:在用户没有电脑或还没上网的时候,WorldUnion 系统可以使他们享受 E-mail 的服务,它把现在的电脑上网发送、接收 E-mail 邮件简化为用电话发送或接收。WorldUnion 系统可以安装在 E-mail 服务器上,在用户说出通话方的电话号码和电话内容后,就近站点服务器上的 WorldUnion 语音识别模块识别电话语音内容,根据电话号码的地理位置传送识别结果。目的站点 E-mail 服务器判别邮件的语种,经翻译模块翻译(或不翻译)成该国语言后合成语音,接通对方录音电话使之接收,反之亦然。在 WorldUnion 系统日渐完善、网络的传输延迟及信元丢失率日益改善的前提下,WorldUnion 电话 E-mail 系统是完全可以实现的。

还有一个关系到普及性和实用性的问题就是系统尺寸,人们不可能提着一个大电脑去交流。随着大规模半导体集成电路技术的高速发展,现在可以在一块芯片上制造含有上千万个晶体管的电路。这样就可能把 CPU、内外存储器和经固化的相关软件集成在一片很小的芯片上,使整个系统尺寸如同一计算器大小,再配上微型耳机话筒,用户即可以自由携带使用。

有理由相信,经过许多科技工作者的共同努力,继续完善系统的各个模块和它们之间的联系,这套系统在不久的将来一定会在世界范围内普及开来。

参考文献:

- [1] 丁晓青,郭繁夏. 汉字识别及其发展和应用[C]. 罗振声,袁毓林. 计算机时代的汉语和汉字研究学术论文集. 北京:清华大学出版社,1996. 317-324.
- [2] 戎月莉. 计算机模糊控制原理及应用[M]. 北京:北京航空航天大学出版社,1995. 164-17.

- [3] [日]神沼二真,松本元. 生物计算机——日本的下一代计算机[C]. 顾仲梅译. 北京:国防工业出版社,1996. 121-135.
- [4] 郑人杰,殷人昆,陶永雷. 实用软件工程[M]. 北京:清华大学出版社,1997. 89-96.
- [5] 谢启江. Windows3. X 应用程序设计技术[M]. 北京:人民邮电大学出版社,1994. 660-735.
- [6] 方棣堂. 汉语语音识别的现状与展望[C]. 罗振声,袁毓林. 计算机时代的汉语和汉字研究学术论文集. 北京:清华大学出版社,1996. 325-336.
- [7] 陈力为. 汉语书面语的分词问题[C]. 罗振声,袁毓林. 计算机时代的汉语和汉字研究学术论文集. 北京:清华大学出版社,1996. 7-10.
- [8] 吴宗济. 为改进合成普通话口语自然度所需韵律特征规则的设计[C]. 罗振声,袁毓林. 计算机时代的汉语和汉字研究学术论文集. 北京:清华大学出版社,1996. 355-365.
- [9] 茅于杭. 为盲人设计的汉字计算机系统[C]. 罗振声,袁毓林. 计算机时代的汉语和汉字研究学术论文集. 北京:清华大学出版社,1996. 414-425.

(编辑 曹大刚)

The study of WorldUnion, a multi-language real-time automatic communication system

LIU Wei, SUN Ya-ru, HE Xiao-yang

(Department of Computer Science Northwest University, Xi'an 710069, China)

Abstract: WorldUnion system gives us a new feasible way to solve the problem that people can not communicate freely when they speak different languages. It makes available software such as speech recognition, language translation and speech synthesis by means of software integration. The main function of the system is to receive and recognize a kind of language speech, then translate this language into another one and synthesize speech. The system is applicable in many fields such as networks communication and commercial cooperation. It is hopeful in many ways.

Key words: speech recognition; language translation; speech synthesis; software integration; DDE

(上接第 207 页)

参考文献:

- [1] Pecora L M, Carroll T L. Synchronization in chaotic systems[J]. Phys. Rev. Lett., 1990, 64(2): 821-824.
- [2] Cuomo K M, Oppenheim A V. Circuit implementation of synchronized chaos with application to communication[J]. IEEE Trans., Circuits syst., 1993, 40(10): 626-633.

(编辑 曹大刚)

The research about chaotic interval synchronization

HUANG Xian-gao¹, XU Jian-xue¹, CHEN Gao-ping²

(1. Institute of Nonlinear Dynamics, Xi'an Jiaotong University, Xi'an 710038, China;

2. Department of Avionics Engineering, The Air Force Institute of Engineering, Xi'an 710038, China)

Abstract: The scheme of interval synchronization is proposed. Synchronizing signal of chaos is automatically extracted from modulated signal by forced synchronization and freely evolving synchronization. Synchronization is maintained between transmitter and receiver under normal modulation ratio. So speech transmitted is exactly recovered. Finally, Computer simulation of Lorenz chaotic system is given.

Key words: forced synchronization; freely evolving synchronization; chaos secure communication