Advances in Language and Literary Studies
ISSN: 2203-4714
www.alls.aiac.org.au

# English voices in 'Text-to-speech tools': representation of English users and their varieties from a World Englishes perspective

Ali Karakaş*

*Department of Foreign Language Education, Mehmet Akif Ersoy University, Burdur, Turkey*

**Corresponding Author:** Ali Karakaş, E-mail: akarakas@mehmetakif.edu.tr

**ARTICLE INFO**

*Article history*
Received: June 30, 2017
Accepted: September 24, 2017
Published: October 31, 2017
Volume: 8 Issue: 5
Advance access: October 2017

Conflicts of interest: Non
Funding: None

**Key words:**
World Englishes,
Text-To-Speech Technology,
English Language Teaching,
Linguistic Diversity,
Concentric Circles,
Lingua Franca,
Language Varieties

**ABSTRACT**

English has experienced grave transformations recently in terms of socio-demographic and geographical characteristics. While such transformations have resulted in diverse types of English uses and various English users, the existing ELT materials still fail to represent the global varieties and dynamic uses and users of English. Moving from a World Englishes perspective, this paper investigates a corpus of online Text-to-Speech tools and software to discuss their suitability for teaching English according to the plurithic view of English, which throws focus on various users and uses of English. Analysed via quantitative content analysis, the data showed that TTS tools promoted the Inner circle (native-English) varieties over the Outer and External circle (non-native) varieties and non-native accents. In addition, the absolute absence of users from the Expanding circle was observed as no speakers from this circle was available in the tools analysed. The findings suggest that a satisfactory World Englishes perspective has not yet been taken into consideration in the present Text-to-Speech tools. There is, thus, a crucial need for a shift in the design of such tools to get them adjusted to represent different types of English users and uses.

## INTRODUCTION

English language teaching (ELT) profession enjoys a long history, which has gradually been shaped by advancements in technology and theories of learning and language. Taught primarily as a foreign language (EFL) in non-English dominant settings as required part of school curricula, English has been going through dramatic transformations for several decades in terms of its speaker profile, areas of use, and functions of its linguistic features. It now has, in statistical terms, more non-native English speakers (NNESs) than its native English speakers (NESs), as noted long ago by some linguists (e.g. Brumfit 2001; Crystal, 2008). It is also dynamically used all around the world in various domains, ranging from education, tourism, aviation, business to politics. Because of changes in socio-demographic and geographic characteristics of English, English has pluralized and taken up the role of a world lingua franca.

### Research Problem and Rationale

The transformations English has undertaken have made it obvious that traditional teaching methods and materials closely aligning with the EFL philosophy will no longer be sufficien in meeting students' linguistic needs as language users. It is because most English speakers have already used and will largely be using English within their national contexts (e.g. to study in an English-medium program alongside many international students and academic staff) and within their workplaces (e.g. in the domains of business, tourism, aviation) predominantly with speakers who do not speak English as their L1. However, in the case of EFL, the aim is to prepare students to chiefly use English with NESs in contexts dominated by their codified norms (Jenkins, 2006). Accordingly, from an EFL perspective, instead of varied users and uses of English, NESs and their ways of doing English are mostly introduced to students since the desired goal for students is set as to reach near-native competence. However, this way of traditional thinking treats deviations in speakers' language use and communicative strategies they employ (e.g. code-switching, L1 use, repairs, and repetitions) as signs of linguistic deficiency and does not match the current reality of English. The main reason for this is that linguistic departures from NES norms and the use of various linguistic, pragmatic and socio-cultural communicative strategies are inevitably inherent in real-life communication among different users of English.

Published by Australian International Academic Centre PTY.LTD.

http://dx.doi.org/10.7575/aiac.alls.v.8n.5p.108

Another reason for why the EFL perspective is irrelevant to the current reality of English is that most learners predominantly use English in non-Anglophone contexts with NNESs and that English is no longer a foreign language to many students. As argued by scholars (e.g. Leung & Street, 2012), practices and linguistic norms associated with NESs and their varieties have lost their relevance to most students. Thus, as Friedrich (2012, p. 50) argued, '[i]f the only constant in lingua franca situations is diversity, then we should anchor our practices in that assumption and educate students to encounter such diversity with respect, curiosity and wisdom'. The reasons cited so far feature the need to innovate the way English is conceptualised, taught and learned. In that sense, the field of World Englishes (WE) has a lot to offer as to innovating the way English is taught, used and the way it should be conceptualised. Scholars in the field of WE have, in this respect, called for more of a focus on particular aspects of the learning and teaching process, including primarily ELT materials (e.g. Matsuda, 2002a, 2002b). They also argued that teaching English in accordance with its socio-linguistic reality falls into the remit of language teachers. One main task assigned to teachers is to raise students' awareness about the fact that English is now being used by a large number of speakers and relatively differently from the predetermined conventions instructed in most language teaching classrooms (McKay, 2012a). It is therefore essential for language teachers to expose students to as many varieties of English as possible (D'Souza, 1999). Both previous WE and ELF, i.e. English as a Lingua Franca, (formerly known as EIL, namely English as an International Language) research have shown that exposure to distinct varieties of English has increased people's familiarity with these varieties and international understanding. Furthermore, research demonstrated that exposures of this kind have positively influenced people's attitudes towards varieties of English and their speakers, leading to an increased comprehension level in communication situations with the speakers of these varieties (e.g. Genç, 2012; Karakaş, 2016; Suviniitty, 2009).

Besides exposing learners to variation in English, teachers are advised to put students in an 'open engagement with differences across uses, users, and contexts of English' so that they can gain a deeper awareness and acceptance of linguistic diversity (Kitazawa, 2012, p. 261). However, the main obstacle teachers encounter in following such an innovated teaching is the lack of teaching materials that can accurately display the non-prescribed varieties and divergent uses of English. This obstacle stems from the fact that most English teaching materials are designed by NESs and according to the particular prescribed varieties of English, often the UK and American English models (Tollefson, 2000).

Seeing the insufficiency of the current ELT materials and tasks in preparing students for real-life communication and meeting their current linguistic needs in the context of pluricentricity of English, many ELF and WE researchers have questioned whether mainstream ELT materials have started to acknowledge the diversity of English uses and users, intercultural and multilingual practices. In this regard, previous studies on ELT materials primarily studied printed materials such as textbooks in various educational contexts, ranging from Japan, Italy, Finland, Germany to Turkey (e.g. Gülle et al., 2016; Matsuda, 2002b; Savalainen, 2012; Syrbe & Rose, 2016; Vettorel & Bayyurt, 2016; Vettorel & Lopriore, 2013) to determine the representation or non-representation of varied users and uses of English. Consistently, the findings of those studies indicated the dominance of the Inner circle varieties and users and the low presence or, sometimes, an absolute absence of others, particularly the Expanding circle and the Outer circle (see Figure 1 below). However, previous studies to date have largely focused on printed teaching materials, ignoring the potential of online materials that can be tailored to be used in language teaching according to the WE paradigm. Consequently, there is currently no research dealing with the case of text-to-speech (TTS) tools although they have been used as a language learning material for at least more than a decade to linguistically support different groups of students (EFL, ESL, ENL [English as a Native Language]) in various minor and major skills.

One can argue that if exposure to different accents, pronunciations and varieties is the only goal, then audio and video materials on the net offer a more accessible solution than the TTS tools. This is not to say that exposure to varieties is not vital because '[a]n incomplete presentation of the English language may also lead to confusion or resistance when students are confronted with different types of English users and uses' (Matsuda, 2002a, p. 438). While audio-video materials can be partly useful for helping students gain some degree of familiarity with English varieties, they are likely to fall short of reflecting the fluid and dynamic use of English by various speakers to a satisfactory extent. TTS tools can, in contrast, be more adaptably used for the same goal and above all, to display how global users of English use English in various real-life communicative domains (e.g. academic/non-academic, formal/informal and leisure) by drawing on the data from the available corpora on various English speakers, especially those from the Expanding and Outer circles. Teachers can, for instance, easily present varied uses and users of English to their students from these corpora by extracting the relevant sections on certain linguistic factors (e.g. lexis, vocabulary, communicative strategies, innovative and creative language use) from any of these corpora and then converting them into audio files through TTS tools.

What also makes the study of TTS technology intriguing is its psychological facet and promise as a tool that can be further improved for language teaching, with more and more naturally sounding and comprehensible voices (Hirai & O'ki, 2011; Kataoka, 2009). Inasmuch as language teaching is a tiresome process requiring a long-term investment and commitment on the part of learners, the integration of TTS technology into language classes can make this process more enjoyable compared to the printed materials like textbooks. While textbooks can be considered a natural starting point for language learners (Pim, 2013), they do not suffice to cater for students' further linguistic and communicative needs in contact situations with different speakers of English. However, despite the possible uses of TTS technology for teaching English in keeping with WE principles, it ap-

pears that the great potential of the TTS tools for teaching English has not been seriously noticed by language researchers hitherto. Consequently, the major purpose of this paper is to critically analyse a large number of freely available online TTS tools and software and then to explore the representation, or lack of representation, of a range of voices, i.e. users and varieties of English in those tools and software from a WE perspective.

### Aim of the Study and Research Questions

Despite the growing prominence of the WE paradigm together with its pedagogical implications for the field of ELT, there are not available materials on WE that teachers, students and scholars can use in teaching. As shown above, there have been calls for changes to ELT practices in consequences of the linguistic and socio-demographic transformations English has undergone. However, it seems that there have not been enough attempts made in practical terms to bring innovations to the pedagogy in teaching English. For this particular reason, this research sees TTS tools as an alternative material to the traditional ones for teaching English in accordance with the characteristics of the WE paradigm. This study, therefore, looks at the current state of the TTS tools in terms of its suitability to be used as an educational material for WE instruction, with a specific intention to answer the following research questions:

(1) To what degree do freely available online TTS programs and software represent global users and uses of English in their voice collections?
   a. in terms of speaker representation based on Kachru's model
   b. in terms of the number of speakers available for a particular variety
(2) To what extent are the current TTS programs and software suitable for teaching English in accordance with the tenets of WE?

## LITERATURE REVIEW

### Text-to-Speech Technology (TTS) and Language Teaching

As one of the recent tools of information technology, TTS technology is regarded as 'a form of speech synthesis that converts text into spoken voice output' (Beal, n.d., para. 1) by means of specifically designed computer programs (Kılıçkaya, 2006). Its origins date back to the 1970s when this technology was basically exercised as an assistive technology to serve individuals with visual impairments and learning difficulties (Baker 2014). However, language teachers and researchers have not noticed the potential of TTS technology for language learning until the beginning of the 2000s due mostly to some sorts of resistance against and lack of attention to such tools. Among many others, take, for example, the shortcomings pinpointed by Higgins (1998, p. vii) who emphasised that '[i]f it cannot account for the full complexity of human language, why even bother modelling more constrained aspects of language use'. Likewise, Garrett (1998, p. 81) alerted that '[t]his technology isn't at a stage where it can reliably render a target language accent authentic enough for language use'. In brief, what they regarded as a downside was that the TTS technology was far cry from generating natural, near human-like voices that can be usefully put into students' service in language classrooms. Their arguments were convincing considering the very first phase of its introduction and the state of the technology at that time.

Following the recent developments in information technology, it is, however, now possible for TTS technology to generate almost near human-like voices. In support of this assertion, several comparative studies on the impacts of TTS voices on listeners demonstrated that listeners did not report a perceptual gap or artificiality between TTS sounds and actual human voices (Hirai & O'ki, 2011; Kataoka, 2009; Jones, Berry & Stevens, 2007). It was even discovered that students with low proficiency level of English preferred TTS sounds to actual NES voices (Hirai & O'ki 2011). Implications of these studies are that TTS tools can be used in language education in general, and more specifically in first language acquisition, second language acquisition, and EFL teaching/learning in different fashions. It is perhaps for this reason that scholars, such as Kılıçkaya (2011), acknowledged its potential for language teaching for a number of purposes (e.g. to work on frequently mispronounced words, to listen to various materials of different genres, and to create dialogues). Additionally, Kılıçkaya suggested that '[l]anguage learners and teachers need to be informed about this technology, its possible uses, advantages and limitations since this technology is new to them' (para. 3). Likewise, with a particular focus on teaching foreign languages, Azuma (2008, p. 498) accentuated that 'the time may have come when we can use the TTS synthesized speech as a model in educational settings focused on the teaching of foreign languages' despite the reality of grave doubts about TTS technology. It is vital to note at this juncture what this technology can actually do in practical terms. Posing this query earlier, Kılıçkaya (2006, para. 11) summarised that TTS tools can

- [r]ead any text in computer (web pages, word documents, rich texts, e-mails, news articles, online books etc.)
- [g]ive the option of reading any text and saving it into a file in the form of wav or mp3 files, which gives the opportunity to listen to them later in your MP3 or CD player.
- [r]ead any text at any speed and any speaking quality
- [r]ead any text using the voice or any accent (male, female, British English, American English, etc.)

Thanks to the above features, the TTS technology is currently being used in several applications, such as desktop speech systems, computer voice interfaces, audio books, and electronic dictionaries (Moon, 2012). Naturally, all these applications can be made part of language education by language teachers. Previous studies, for instance, indicated that primary school students in Canada practising reading in their L1 through a TTS program made a significant progress in improving their reading skills (Parr, 2013). The TTS applications were also found to be practical for ESL students, espe-

cially those with low literacy skills in increasing their reading performance (Baker, 2014; D'Silva, 2005). Likewise, research showed the benefits of these tools for NNES students as to developing their reading comprehension (Drezek, 2007) and making progress at word recognition, reading, writing and spelling, and pronunciation (Chiang & Liu, 2011), as well as vocabulary memorizing (Kataoka, 2007). Additionally, researchers observed the positive impact of using audio materials developed via TTS voices in enhanced listening comprehension (Sha, 2009).

Added to the above benefits, scholars, taking a conventional SLA stance on this issue, have contended that the TTS tools are capable of exposing learners to meaningful and comprehensible input through listening exercises, especially in input-poor educational contexts where English is barely used/spoken outside the official teaching situations. Regarding this matter, some researchers, such as Moon (2012, p. 120), maintained that TTS tools can provide 'great potential for offering learners with varied and easily accessible spoken language input' and hence 'can add variations to listening comprehension using different voices'. While giving examples of different voices, he, however, accentuated the point that '[m]any of these programs allow users to select voice and accent ranging from female to male and American to British' English' (Moon 2012, p. 121). In a sense, he contradicted himself while pointing out the benefits of TTS tools as regards to creating varied and diverse voices. It is because, to him, varied and different voices in English seem to be restricted to British and American Englishes only.

Adopting a similar line of thought, Azuma (2008, p. 498) has concluded that language teachers and learners can use 'a lot of variety of TTS generated voices' in any language they prefer, yet primarily in demographically and institutionally dominant languages, such as English and Spanish, and more specificall in their standard versions. Earlier, a similar view was articulated by Kılıçkaya (2006, para. 11) stating that TTS programs can '[r]ead any text using the voice or any accent'. Nonetheless, he overtly alluded to British and American English voices only when exemplifying the options teachers and learners have at their disposal, and did not mention the possibility of using other English voices available, particularly those of NNESs. Moreover, he tested TTS technology in experimental studies with EFL learners for the purposes of accent reduction, namely, to help students eliminate their own L1 traits from the way they use English (e.g. Kılıçkaya, 2008, 2011). Although Kılıçkaya himself acknowledged the triviality of targeting NES competence in the teaching of pronunciation, suggesting an intelligibility benchmark, he recommended TTS tools for accent reduction, self-contradicting his intelligibility argument regarding pronunciation. As widely known, accent reduction sessions are intended for individuals with foreign-accented speech so that they can remove their accents and sound just like a NES as much as possible by giving up their own ways of using English.

From the above arguments and research findings of the previous studies into the TTS technology, it has become evident that TTS technology has mainly been considered in language education from a traditional SLA and an EFL perspective and that previous studies have not associated TTS technology with WE so far. It can also be presumed that most researchers viewed TTS technology as a great tool for learners' exposure to so-called authentic and natural input at the absence of NESs. However, if used in accordance with WE principles, TTS tools can enable linguistic diversity to be acknowledged and better understood by students.

## World Englishes Research Paradigm

World Englishes, as a research field, has emerged as a reaction to its former precedents, i.e. ENL, SLA and EFL models along with their inherently problematic and outdated notions about language, e.g. interlanguage, native speaker/non-native speaker divide, fossilization, and deficit view of language learners (Brutt-Griffler & Samimy, 2001; Kachru, 1986; McArthur, 1993). WE scholars have severely problematized the deficit approach to NNESs and the outdated and conformist conceptualisations of English. Moreover, they did not remain indifferent to the theoretical and pedagogical implications of the transformations English has been experiencing over the years in terms of the increased number of speakers and areas of use, and constantly changing communicative needs of its speakers. Therefore, the WE paradigm has adopted a plurithic view of English, taking into account its wide spread across the world and primarily dealt with non-native ways of doing English. A plurithic view of this kind 'represents diverse sociolinguistic histories, multicultural identities, multiple norms of use and acquisition, and distinct contexts of function' (Bhatt, 2001, p. 527). Drawing on this view of English, WE scholars have rejected the earlier models noted above and embraced 'a model of diffusion of English that is defined with reference to historical, sociolinguistic, and literary contexts' (Bhatt, 2001, p. 528). Moreover, the WE paradigm has attempted to liberate English from the ownership of NESs and their corresponding norms, supporting the ownership of it by its all users all around the world. In this regard, some scholars harshly criticized the native and non-native divide, supporting the contention that '[n]ational identity should not be a basis of classification of speakers of an international language' like English (Brutt-Griffler & Samim , 2001, p. 105).

WE, as a research field, studies 'varieties of English used in diverse sociolinguistic contexts' (Bhatt, 2001, p. 527), and is primarily concerned with codification of national varieties. To better study varieties of English, several models of the spread of English as a global language were constructed by WE scholars, the most influential of which is Kachru's (1986, 1992) concentric circles, i.e. a theoretical framework developed by himself to clearly display the pluricentricity in English and its speakers around the world. This framework is composed of three circles: The Inner circle, Outer circle and Expanding circle (see Figure 1 below). According to this division, the Inner circle countries are the long-established bases of English (e.g. UK, the USA, Canada) where English is primarily spoken as a mother tongue by its native speakers. The Outer circle countries are those (e.g. India, Zambia) where English functions as an L2 due to its being institu-

tionalized as an additional language and is thus often used for intranational communication. Finally, the Expanding Circle includes countries where English has not been institutionalized yet, and is thus mainly used for international but less for intranational communication. English is essentially taught/learned as a foreign language in schools of the countries belonging in this circle. However, English has started to be used for intranational functions in the Expanding circle countries recently, mostly in educational domains (e.g. as a medium of instruction), yet such instrumental uses of English are still relatively restricted to very few domains.

The WE paradigm demands instructional changes in ELT practices because the existing practices are in no state to teach English in accordance with the current plurithic face of English. For English to be taught in accordance with the WE paradigm, it is necessary, as proposed by some scholars, that language learners should be made aware of the cultural and linguistic diversity of English and its globalized status, e.g. global uses and users (Matsuda & Friedrich, 2012). Moreover, linguistic respect should be instilled within learners and teachers so that they can show tolerance towards the linguistic divergences of their future interlocutors whose use of English does not fit the prescribed norms of standard English taught at school contexts. For this, they should be informed about the multilingual nature of English, along with the individual strands relating to multilingualism like code-switching and translanguaging (Dewey, 2012; Jenkins, 2015; McKay, 2012b). For all these to happen, the first step that needs to be taken in the right direction is to raise language learners, teachers, and users' awareness about the plurithic view and diversity of English. It seems that TTS tools have the makings of an alternative tool to the traditional ones in terms of raising learners and teachers' awareness of the current face of English and familiarising them with diverse English voices and ways of using English.

## THE STUDY

### Materials

A corpus of 50 freely available online TTS tools and software was collected and analysed for the purpose of this study. Utmost care was taken to include all the available online TTS tools and software at the time of compilation. All these tools and software can be accessed via the Internet, and some can even be downloaded into any computerized devices for offline use (see Appendix for the list of the TTS tools and software analysed). The TTS tools were chosen on the basis of purposive sampling (Cohen et al., 2007) in which a thorough search of the Internet via the search engine, Google, was conducted with the keywords *text-to-speech tools*, *text-to-speech software*, and *TTS service*. There were only two criteria for a TTS tool to be involved in the list: (i) it should be freely available on the Internet, and (ii) it should be able to generate voices in the English language. The TTS tools meeting these two criteria were included in the corpus for the main data analysis. Most of these online TTS tools are available at users' service for free, while some others are produced for commercial purposes; the users thereby need to pay a certain amount of fee for the use of full versions of the tools. Only the demo versions of the fee-charging TTS tools were analysed in this research. However, this case did not have much influence on the actual data analysis because the demo versions also show which varieties of English are represented in the available voices, yet the generated voices of English cannot be played and listened to in some demo versions.

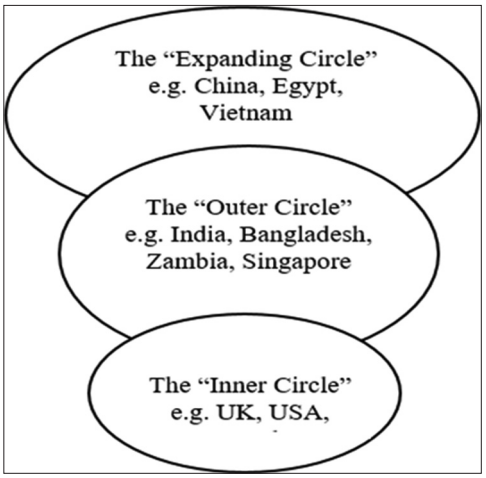


**Figure 1.** Kachru's three-circle model of English

### Method

The analysis of the selected TTS tools was carried out in order to detect the occurrences or non-occurrences of English varieties in the available voices of the TTS tools based on Kachru's three-circle model of English speakers. At some points, the Microsoft Word and Excel were utilized to do some basic descriptive statistics like calculating the frequency of a particular English voice across all the TTS tools analysed. A statistical software, SPSS 22 was partially used in an attempt to run some basic descriptive statistics and create graphs that will nicely display the representation of – lack thereof – the varieties of English in the voices offered by the TTS tools.

For systematic and objective identification and representation of the English varieties presented in the TTS tools, the data were analysed via quantitative content analysis, which aims for "the systematic, objective, quantitative analysis of message characteristics" (Neuendorf, 2002, p. 1). The message characteristics refer, in this research, to the covert promotion and demotion of particular voices of English varieties over others and availability and meaningful absence of certain English voices. The policy of inclusion and exclusion of particular voices evidently conveys a covert message to the users of TTS tools and software as regards their importance and status. Through quantitative content analysis, the aim was to identify the speaker patterns available in the voice repertoire of the TTS tools to see whether English voices across the world are equally represented in these tools or, they are lopsided towards particular voices that represent certain English varieties and their speakers. In other words, the quantitative content analysis was used to account for the

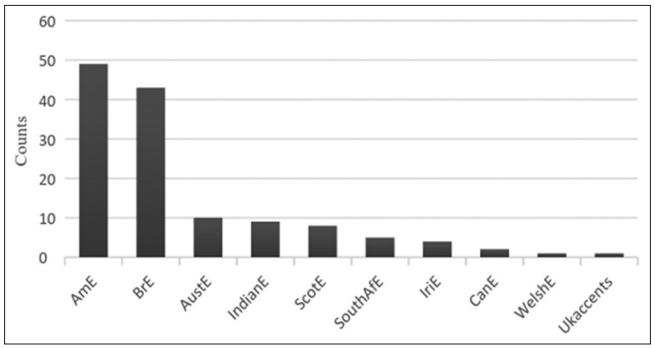


**Figure 2.** Counts of the English voices across 50 TTS tools analysed

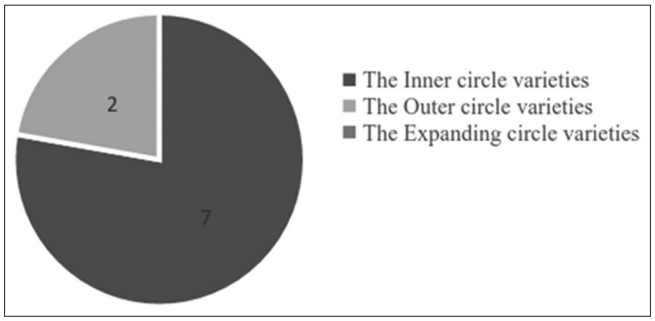


**Figure 3.** Distribution of the speaker voices according to Kachru's three-circle model

frequency of English speakers and voices available in the TTS tools under investigation. To this end, the first thing done was to explore how frequently a particular variety is represented by the TTS tools, of course, if available among the voices offered, and, second, which voices of the English varieties are meaningfully absent from the voice selection box.

The analytical process consisted of a few stages: first, a comprehensive internet search was made to identify the probable TTS tool to be included in the corpus. Second, the identified tools were carefully read and analysed so that English voices/speakers can be classified according to Kachru's three-circle model. Third, the data were transported into various software, like Microsoft Word/Excel and SPSS to run some basic analysis and create a visual representation of the voices/speakers. The occurrences of any English voices and speakers were counted and the percentages and statistical data were calculated.

## RESULTS

In the first part of the analysis, the corpus was primarily scrutinized for the representation of English varieties in their voice collection. With respect to the research question 1, which sought to find out the extent to which WE varieties are represented in the TTS tools, the analysis highlighted that American English (AmE) and British English (BrE) are the most widely represented varieties across the 50 TTS tools as shown in the following fi ure.

As can be seen above, out of 50 TTS tools, 98% (n= 49) offer AmE voices, 45,5 % (n=43) offer BrE voices to their

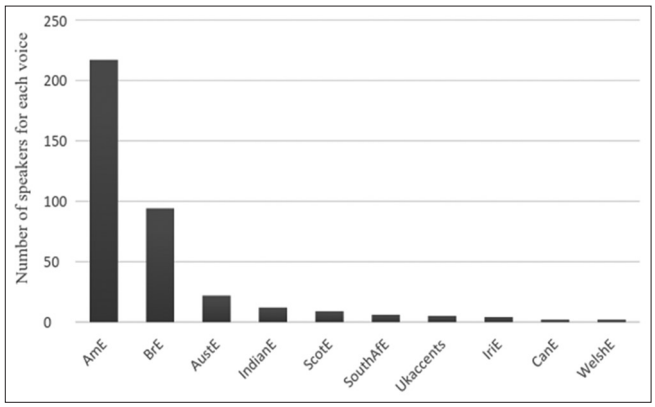


**Figure 4.** Distribution of the number of speakers for the given varieties in TTS tools

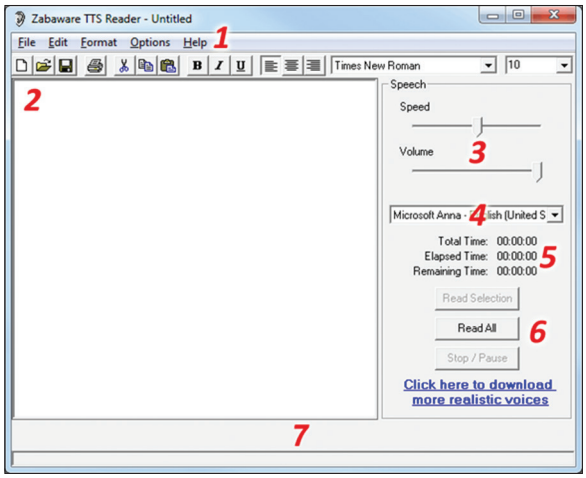


**Figure 5.** Screenshot of Zabaware TTS reader's interface

users. What is striking is that Indian English, albeit being an Outer circle variety, takes precedence of some Inner circle varieties, such as Scottish English, Irish English, and Welsh English on account of voice representation, which are only included in very few TTS tools' voice collections. However, it is also likely that some TTS tools use BrE as an inclusive term, and that a person from Glasgow, for example, is supposed to represent this BrE. Thus, the less dominant Inner circle varieties, like Scottish English and Irish English, can be included in the voice collections, but under the name of BrE, therefore can remain invisible to the users of the TTS tools.

One unanticipated finding was that as shown in the graph below, none of the TTS tools analysed offer voices representing the Expanding circle varieties and speakers. This case is quite startling given the fact that the vast majority of the English speakers belong to this circle. For instance, if one considers the current finding from a WE perspective, a dialog between, say, one South Korean and one German talking with each other mainly through English and with departures from standard English cannot be presented to the learners and users due to the lack of their voices.

As regards the number of speakers playing the voices of the TTS tools, it was found that AmE has the highest number of speakers (see Figure 4 below), followed by BrE. That is, a

user is given the opportunity to listen to AmE and BrE voices from a wide range of male and female speakers.

Any user who wishes to listen to less dominant and subordinate Inner circle varieties, like Canadian English and Welsh English, have only two options: they can either listen to an Irish and a Canadian female voice or an Irish and a Canadian male voice. The users have no other options when they dislike the generated speech or are unsatisfied with the voice quality of the available speakers. However, the most unfortunate are those who make an attempt to generate voices in NNES accents, which are in no way provided by any of the TTS tools analysed. The findings also show that the notion of linguistic diversity is limited to the speakers who represent regional varieties and accents of standard (native) English and to some extent the Outer circle varieties as in the case of Indian English and South African English. There was actually just one tool that explicitly introduced UK regional accents like that of Lancashire. The others did not overtly specify whether the speakers represent different regional accents, yet when the generated speeches were downloaded and played using a music player, it turned out that the speakers' speech markedly differed from each other's. This difference shows that variation only applies to the Inner circle varieties and the Outer circle varieties to a very small extent, but it does not apply to the Expanding Circle Englishes in any place.

Additionally, when the data was closely inspected, it was noticed that the online TTS tools allocate more room for various speakers while the TTS software basically uses the computer's default voice(s) from the Windows language pack in which supported English varieties are restricted to AmE and BrE solely (see, for example, the following screenshot of a TTS tool; number 4 for language choice).

For the reason cited above, while calculating the number of English speakers for a particular variety, two speakers were allocated to each variety, considering that one will be a male, and the other will be a female speaker of that voice.

## DISCUSSION

The results have shown that the TTS tools popularize the Inner circle varieties of English, out of which AmE and BrE overwhelmingly come to the forefront as the most prevalent varieties of English. This is why, the Outer circle and especially the Expanding circle countries are not given equal attention in the voice representation. This finding is in accord with previous studies on the textbooks, indicating that very little space is allocated to the different varieties and users of English, while AmE and BrE receive the most amount of attention and representation (Gülle et al., 2016; Matsuda, 2002a; Vettorel & Bayyurt 2016; Vettorel & Lopriore, 2013).

It also emerged in the findings that there is a hierarchical order of the native English varieties where non-dominant varieties, such as Irish English and Canadian English, are offered to the users to a very small extent. The underrepresentation of varieties such as Scottish English, Irish English, and Welsh English can be explained by the fact that AmE and BrE have already established their places in various domains as the most widely known and the most desired varieties compared to the other Inner circle varieties (Karakaş, 2016). Therefore, the TTS designers might believe that the customers' demand for other inner circle varieties would not be as strong as the demand for AmE and BrE due to the perceived status differences. For example, AmE and BrE are unquestioningly associated with the notions of correctness, prestige, authority and prescriptiveness, but other inner circle varieties are not. There is also the intelligibility issue. Research has indicated that AmE and BrE are often perceived to be easier to understand compared to other native English varieties and non-native English accents (Jenkins, 2007; Lee, 2012; Lippi-Green, 2012; Rogerson-Revell, 2007). For all the reasons cited, the TTS designers might have brushed aside the less well-known and low-status inner and outer varieties, reasoning that it is not worth including such varieties perceived with a low-status in their voice collections for commercial reasons.

One might also wonder why the designers of the TTS tools have chosen to include only two Outer circle varieties, i.e. Indian English and South African English over the others. Although the reason for this is not so clear, it may have something to do with the fact that Indian English and South African Englishes are among the first 'institutionalized non-variet[ies] of English' (Kachru, 1986, p. 92) and that the English language penetrated into those lands far earlier and gained the status of being, in most cases, the only functional lingua franca of these countries. It might also have something to do with the demographic trends of these countries where there is a considerably high number of English speakers using English on a daily basis and in their own ways.

Another remarkable finding was the non-representation of the Expanding circle speakers and varieties in the TTS tools. This complicated situation might be explained by the fact that native-speakerism, as a pervasive ideology (Holliday, 2006) is maintained by the designers of the TTS tools, even if they are not aware of this ideology by its name. Probably urged by this ideology, they might consider the English varieties from the Inner circle countries to be the ideal ones to be made available in their voice collections. Another possible explanation might be that because these TTS tools are originally designed for commercial purposes rather than language teaching/learning, the designers can conceive that the customers would prefer standard (native) English voices to non-standard voices and non-native English accents. Therefore, they may intentionally have left out voices of NNESs and less-known NESs from their voice collections.

Based on the findings, one can argue that the current TTS tools can only be used to a small extent for the WE instruction for some particular reasons. First of all, in contrast to earlier discussions by researchers, such as Kılıçkaya (2006), Azuma (2008) and Moon (2012), who proposed that TTS tools can enable users to produce speech in various English varieties, this study has been unable to substantiate their claim. The main reason for this is that the TTS tools analysed in this research merely allow their users to produce speech in a limited number of English varieties, predominantly the Inner Ci.k vcxz,rcle ones. Hence, for example, a user who would like to produce and listen to a speech with a German

accent or Polish accent cannot achieve this with the existing TTS tools.

Apart from that, it was also argued by researchers (Kılıçkaya, 2011) that language teachers and learners can create their own dialogues that can be used for language learning purposes (e.g. to practice writing, listening, reading). However, in respect to the form of communication, the current TTS tools appear to be more suited for monologic speech in comparison to dialogic and polyadic interactions. The fundamental reason for this is that users can convert texts into speech using a single voice only even if the text is made up of a dialog or a group talk. That is, a dialog between two speakers (e.g. an Inner circle speaker and an Outer circle speaker) can be created in the voiceover of a particular variety (e.g. AmE, BrE, Indian English). Thus, in their current state, TTS tools fail to reflect the pluricentricity in English and the true nature of real-life communication in which each individual has their own voice characteristics (e.g. accent, pitch, stress, L1 influence, and regional influence). Another issue with TTS tools is the continuing limitations of TTS tools in the areas of intonation and suprasegmental features. However, counter evidence against this issue was submitted by Kataoka (2009), Hirai and O'ki (2011) in Japan and Jones et al. (2007) in Australia. These researchers uncovered that EFL students listening to sentences and dialogs generated via TTS tools reported that the voices sounded natural and comprehensible. However, their studies included NES voices only; therefore, voices of NNESs can result in a perception gap on the part of listeners between actual human voices and the TTS speech sounds in English education.

## CONCLUSIONS

In this investigation, the purpose was to explore the freely available online TTS engines and software from a WE perspective. Part of the purpose was to account for the English voices and varieties offered by the TTS tools and their distribution according to the concentric circle model of WE speakers. By virtue of doing so, this paper has discussed whether the TTS tools in their current forms can be suitably used as an educational material in the teaching of English in accordance with the main characteristics of the WE paradigm. The study has identified that overall, the TTS tools are skewed towards the Inner circle varieties and their speakers, predominantly AmE and BrE speakers because they offer a large number of Inner circle voices to their users. Nevertheless, once it comes to the representation of the Outer circle varieties and speakers, they offer only Indian English and South African English voices. As for the Expanding circle varieties and speakers, there is not even a single voice offered to the users. These findings have led to the conclusion that the TTS tools cannot be effectively used to expose students to the global use of English at present to a large and satisfactory extent. They might be partially made part of the teaching courses in order to make students/learners familiar with different Inner circle varieties and divergences (e.g. regional accents) within a single native English variety and a couple of Outer circle varieties. However, it is very probable that NNES learners will use English more with NNESs than NESs, particularly in non-English speaking environments, yet the current TTS tools are unable to create platforms in which users can gain an awareness of the diversity of English speakers and the diverse ways of using English, which is an essential part of the WE communication.

These findings, when taken together, have some pedagogical implications which can be born in mind by language teachers, information technology experts, and TTS tool designers. For example, the findings suggest that the TTS tools have a great potential to be used as a language teaching/learning material by language teachers. Particularly, they may play a significant role as a source of exposure to various uses and users of English. However, language teachers should exercise some degree of caution before introducing the TTS tools to language classrooms owing to the fact that several major and specific amendments need to be made to the TTS tools currently in use. As mentioned above, although the TTS tools were not originally released for the purpose of language teaching, they are being used in the language classrooms by teachers for various purposes. However, they are, as the results indicate, presently far from being a relatively effective learning/teaching resource for WE-friendly instruction. In other words, they are not quite fit for purpose at present. Still, language experts and information technology experts can, by collaborating together, bring TTS tools in conformity with the main characteristics of the WE language teaching (e.g. by adding diverse speakers into the voice collections of the tools from all circles; by improving the tools to be appropriate for dialogic and polyadic interactions). With the fast advancing information technology, fine-tuning these tools for the purpose of language education will not be so difficult for the designers, especially when informed by linguists aware of the WE field

It should also be noted that this study has its limitations relative to the investigation of the online TTS engines, and TTS software which can be downloaded and installed to be used offline. First, the analysis has remained limited to 50 TTS tools, most of which are online TTS engines. Although an exhaustive search of the Internet was done to identify the freely available TTS tools and software, there may be some tools missed out from the analysis, which might, however, have offered English voices different from the ones included in this paper. Another limitation of this study is that while accounting for the representation of the English varieties and the number of speakers for each variety, little attention was paid to the representation of social variables, like gender and age group, although such variables have characterized speakers in several TTS tools. Gender equity is not only an important focus of sociolinguistics and the WE paradigm (Yılmaz-Öztürk, 2016) but also, as a variable, gender adds variation to the voice selections of the TTS tools. Some TTS users, for instance, may prefer to generate speeches in female voices or vice versa. The representation of the speakers from different age groups is rather vital, as well. It is because different speakers, ranging from children, adults, to the elderly, realize the global uses of English. Lastly, the investigation in this paper remains restricted to the English voices only.

The abovementioned limitations are in need of further investigation, and can be overcome by researchers in future research. It would be interesting, for example, to explore the same TTS tools in terms of gender and age group variables. A further study can also investigate TTS tools in relation to the teaching of particular sub-skills, such as pronunciation and accent recognition. For this, researchers can even conduct experimental studies, with one group working on a given language skill with TTS tools and the other without TTS tools. Due to not working with human participants, this research is unable to present any information regarding language teachers and learners' perspectives about the potential of the TTS tools for teaching English in line with its current face. Hence, further research is required to explore language teachers and learners' attitudinal views as regards the use of TTS tools as an educational material and its capacities for teaching English as a global language. Last of all, the exploration of c781,40TTS tools in relation to the voices of other international languages, like Spanish and Arabic, can be an important issue for future research.

## REFERENCES


Azuma, J. (2008). Applying TTS technology to foreign language teaching. In F. Zhang & B. Barber (eds.), *Handbook of research on computer-enhanced language acquisition and learning*, (pp. 497-506). New York: Information Science Reference.

Baker, F. S. (2014). Text-to-speech software as assistive and mainstream technology: Transitioning from a functional to a socio-constructivist approach. In B. DaCosta & S. Seok (eds.), *Assistive technology research, practice, and theory*, (pp. 27- 43). Hersley, PA: IGI Global.

Beal, V. (n.d.). *TTS- text to speech*. Retrieved from http://www.webopedia.com/TERM/T/TTS.html

Bhatt, R. M. (2001). World Englishes. *Annual Review of Anthropology*, 30, 527–550.

Brumfit, C. (2001). *Individual freedom in language teaching*. Oxford: Oxford University Press.

Brutt-Griffle , J., & Samimy, K. K. (2001). Transcending the nativeness paradigm. *World Englishes, 20*(1), 99–106. https://doi.org/10.1111/1467-971X.00199

Chiang, H. & Liu, C. (2011). Evaluation of the benefits of assistive reading software: Perceptions of high school students with learning disabilities. *Assistive Technology, 4*(2), 106-118.

Cohen, L., Lawrence, M. & Keith, M. (2007). *Research methods in education* (6th ed.). London: Routledge/ Falmer.

D'Silva, R. A. (2005). *Promoting reading skills of young adult EAL learners through voice recognition software*. (Unpublished master's dissertation). University of British Columbia, Vancouver, Canada.

D'Souza, J. (1999). Afterword. *World Englishes, 18*(2), 271–274.

Dewey, M. (2012). Beyond labels and categories in English language teaching: Critical reflections on popular conceptualisations. In C. Leung & B. Street (eds.). *English – A changing medium for schooling*, (pp. 129-149). Bristol: Multilingual Matters.

Drezek, J. (2007). *Adult ESOL reading comprehension and text-to-speech software*. (Unpublished master's dissertation). University of British Columbia, Vancouver, Canada.

Garrett, N. (1991). Technology in the service of language learning: Trends and issues. *The Modern Language Journal, 75*(1), 74-101.

Genç, Z. S. (2012). May. *Learners' perceptions of ELF: The Turkish context*. Paper presented at the Fifth International Conference of English as a Lingua Franca. Boğaziçi University, Istanbul, Turkey.

Gülle, T., Kurt, Y. & Sönmez, E. (2016). *Representation of English in ELT textbooks: Any room for EIL/ELF?*. The 9th International Conference of English as a Lingua Franca, Book of abstracts, 40. Retrieved from http://www.elf9.org/wp-content/uploads/2016/06/ELF_9_Programme_and_book_of_abstracts.pdf

Higgins, J. (1988). *Language, learners and computers*. London: Longman.

Hirai, A. & O'ki, T. (2011). Comprehensibility and naturalness of Text-To-Speech synthetic materials for EFL listeners. *JACET Journal,* 53, 1-17

Holliday, A. (2006). Native-speakerism. *ELT Journal, 60*(4), 385–387.

Jenkins, J. (2006). Points of view and blind spots: ELF and SLA. *International Journal of Applied Linguistics, 16*(2), 137–162.

Jenkins, J. (2007). *English as a lingua franca: Attitude and identity.* Oxford: Oxford University Press.

Jenkins, J. (2015). Repositioning English and multilingualism in English as a Lingua Franca. *Englishes in Practice, 2*(3), 49–85.

Jones, C., Berry, L. & Stevens, C. (2007). Synthesized speech intelligibility and persuasion: Speech rate and non-native listeners. *Computer Speech and Language, 21*(4), 641-651.

Kachru, B. B. (1986). *The Alchemy of English: The spread, functions and models of non-Native Englishes.* London: Pergamon.

Kachru, B. B. (1992). *The Other tongue: English across cultures* (2nd ed.). Urbana: University of Illinois Press.

Karakaş, A. (2016). *Turkish lecturers' and students' perceptions of English in English-medium universities*. (PhD dissertation). Southampton: Southampton University.

Kataoka, H. (2007, December). *Preparation for university entrance examinations: how to learn frequently-appearing English words with the TTS audios*. Paper presented at the Multimedia & Internet Seminar, Japan Association for Language Education and Technology (LET), Kansai Chapter, Osaka, Japan.

Kataoka, H. (2009). The use of Text-To-Speech (TTS) synthesis technology for English education: Speech recognition of Japanese EFL learners [Text-To-Speech (TTS) synthesis technology o katsuyoushita eigo kyouikukyouzai no kaihatsu to nihonjin no onseininshiki]. *Journal of Kansai University Graduate School of Foreign Language Education and Research,* 7, 1-33.

Kılıçkaya, F. (2006). Text-To-Speech Technology: What Does It Offer To Foreign Language Learners?. *CALL-EJ Online, 7*(2). [Online]

Kılıçkaya, F. (2008). Improving pronunciation via accent reduction and text-to-speech software. In T. Koyama, J. Noguchi, Y. Yoshinari & A. Iwasaki (eds.). Proceedings of the WorldCALL 2008 Conference, Japan, 1, 135–137.

Kılıçkaya, F. (2011). Improving pronunciation via accent reduction and text-to-speech software. In M. Levy, F. Blin, C. B. Siskin & O. Takeuchi (eds.), *WorldCALL: International perspectives on computer-assisted language learning*, (pp. 85-96). New York, NY: Routledge.

Lee, H. (2012). An investigation of Taiwanese university students' attitudes towards ELF. (Unpublished doctoral dissertation). Southampton University, Southampton, UK.

Lippi-Green, R. (2012). *English with an Accent: Language, Ideology, and Discrimination* (2nd ed.) London: Routledge.

Matsuda, A. & Friedrich, P. (2012). Selecting an Instructional Variety for an EIL 56 Curriculum. In A. Matsuda. (ed.). *Principles and practices of teaching English as an international language*, (pp. 17-26). Bristol: Multilingual Matters.

Matsuda, A. (2002a). "International understanding" through teaching world Englishes. *World Englishes*, *21*(3), 436–440. http://doi.org/10.1111/1467-971X.00262

Matsuda, A. (2002b). Representation of users and uses of English in beginning Japanese EFL textbooks. *JALT Journal, 24*(2), 49-62.

McArthur T. (1993). The English language or the English languages? In W. F. Bolton & D. Crystal (eds.), *The English Language* (pp. 323– 341). London: Penguin Books.

Mckay, L. M. (2012b). Principles of teaching English as an international language. In L. Alsagoff, G. Hu, S. L. Mckay & W. A. Renandya (eds.), *Principles and Practices for Teaching English as an International Language*, (pp. 28–46). New York, NY: Routledge.

McKay, S. L. (2012a). Teaching materials for English as an international language. In A. Matsuda (ed.), *Principles and practices of teaching English as an international language*, (pp. 70–83). Bristol: Multilingual Matters.

Moon, D. (2012). Web-Based Text-to-Speech Technologies in Foreign Language Learning: Opportunities and Challenges. In T. Kim, J. Ma & W. Fang (eds), *Computer Applications for Database, Education, and Ubiquitous Computing*, (pp. 120-125). Springer Berlin Heidelberg.

Neuendorf, K. A. (2002). *The content analysis guidebook*. Thousand Oaks, CA: Sage.

Parr, M. (2013). Text-to-speech technology as inclusive reading practice: Changing perspectives, overcoming barriers. *LEARNing spaces, 6*(2), 303-322.

Pim, C. (2013). Emerging technologies, emerging minds: digital innovations within the primary sector. In G. Motteram (ed), *Innovations in learning technologies for English language teaching*, (pp. 17-42). UK: Spring Gardens

Rogerson-Revell, P. (2007). Using English for international business: A European case study. *English for specific purposes, 26*(1), 103-120.

Sha, G. (2009). Using TTS voices to develop audio materials for listening comprehension: A digital approach. *British Journal of Educational Technology, 41*(4), 632-641.

Suviniitty, J. (2009, July). What you have to understand is. – Interactional features in lectures. Paper presented at *the SEFI the Société Européenne pour la Formation des Ingénieurs* – European Society for Engineering Education. Rotterdam.

Syrbe, M. & Rose, H. (2016). An evaluation of the global orientation of English textbooks in Germany. *Innovation in Language Learning and Teaching,* 1-12.

Tollefson, J. W. (2000). Policy and ideology in the spread of English. In J. K. Hall & W. Eggington (eds), *The sociopolitics of English language teaching,* (pp. 7-21). Buffalo: Multilingual Matters Ltd.

Vettorel, P & Lopriore, L. (2013). Is there ELF in ELT coursebooks?. *Studies in Second Language Learning and Teaching, 3*(4), 483-504.

Vettorel, P. & Bayyurt, Y. (*2016). ELF-aware teaching materials: the case of Turkey and Italy*. The 9th International Conference of English as a Lingua Franca, Book of abstracts, 68.

Yılmaz-Öztürk, S. (2016). *Gender representation in EFL textbooks used at secondary schools in Turkey*. The 9th International Conference of English as a Lingua Franca, Book of abstracts, 70.

**Appendix: List of the TTS tools analysed**

| Type of the TTS tool | Name | Available from |
|---|---|---|
| 1. Online | Text to Speech | http://www.fromtexttospeech.com/ |
| 2. Online | Read Speaker | http://www.readspeaker.com/voice-demo/ |
| 3. Online | Acapela box | https://acapela-box.com/AcaBox/index.php |
| 4. Online | Oddcast | http://www.oddcast.com/home/demos/tts/tts_example.php?sitepal |
| 5. Online | TTS Reader | http://ttsreader.com/ |
| 6. Online | Natural reader | http://www.naturalreaders.com/index.html |
| 7. Online | Text2Speech | http://www.text2speech.org/ |
| 8. Online | Ivona | https://www.ivona.com/ |
| 9. Online | Readthewords | http://www.readthewords.com/Try.aspx |
| 10. Online | Spoken Text | http://www.spokentext.net/ |
| 11. Online | Text-to-speech | http://text-to-speech.imtranslator.net/ |
| 12. Online | TTS by iSpeech | http://www.ispeech.org/text.to.speech |
| 13. Online | Yakitome | https://www.yakitome.com/tts/text_to_speech/Audrey?b=536966 |
| 14. Online | Cepstral | http://www.cepstral.com/en/demos |
| 15. Online | Code Welt | http://codewelt.com/proj/speak |
| 16. Online | vozMe | http://vozme.com/index.php?lang=en |
| 17. Online | TTS Online | http://tts.softgateon.net/ |
| 18. Online | Responsive voice | http://responsivevoice.org/ |
| 19. Online | Neo Speech | http://neospeech.com/ |
| 20. Online | Voice Forge | http://www.voiceforge.com/demo |
| 21. Online | Text2speech | http://text2speech.us/ |
| 22. Online | Linguatec | http://www.linguatec.net/products/tts/voice_reader/vrs15demo |
| 23. Online | Wizzard | http://wizzardsoftware.com/text-to-speech-sdk.php |
| 24. Online | Lumenvox | http://www.lumenvox.com/products/tts/#chooseLanguage |
| 25. Online | Sitepal | http://www.sitepal.com/text-to-speech/ |
| 26. Online | Spoken Text | https://www.spokentext.net/ |
| 27. Online | Vocalizer 6 | http://www.nuance.com/landing-pages/playground/Vocalizer_Demo2/vocalizer_modal.html?demo=true |
| 28. Software | Balabolka | http://www.cross-plus-a.com/balabolka.htm |
| 29. Software | TTS Maker | http://downloads.tomsguide.com/Text-Speech-Maker, 0301-5741.html |
| 30. Software | Zabaware | https://www.zabaware.com/reader/ |
| 31. Software | AnalogX Sayit | http://www.freewarefiles.com/AnalogX-Sayit-V_program_583.html |
| 32. Software | DSpeech | http://www.freewarefiles.com/DSpeech_program_19529.html |
| 33. Software | SayPad | http://www.freewarefiles.com/SayPad_program_70044.html |
| 34. Software | Read This | http://www.freewarefiles.com/Read-This_program_67731.html |
| 35. Software | ClipSpeak | http://www.freewarefiles.com/ClipSpeak_program_42972.html |
| 36. Software | Language Reader | http://www.freewarefiles.com/Language-Reader_program_19573.html |
| 37. Software | TTSREader | http://www.freewarefiles.com/TTSReader_program_42660.html |
| 38. Software | Text2Speech | http://www.freewarefiles.com/TextSpeech_program_41173.html |
| 39. Software | HearPC | http://www.freewarefiles.com/HearPC_program_35946.html |
| 40. Software | SmartRead | http://www.freewarefiles.com/SmartRead-Build_program_13754.html |
| 41. Software | Speak Text | http://www.freewarefiles.com/Speak-Text_program_20353.html |
| 42. Software | Speak Clipboard | http://downloads.fyxm.net/Speak-Clipboard-11908.html |
| 43. Software | Word Talk | http://www.wordtalk.org.uk/home/ |
| 44. Software | Cool Speech | http://download.cnet.com/CoolSpeech/3000-33660_4-75439901.html |

(*Contd...*)

**Appendix:** ***(Continued)***

| Type of the TTS tool | Name | Available from |
|---|---|---|
| 45. Software | Panopreter | http://download.cnet.com/Panopreter-Basic/3000-33660_4-10758886.html |
| 46. Software | MWS Reader | http://download.cnet.com/MWS-Reader/3000-33660_4-75998615.html |
| 47. Software | ToVoice | http://download.cnet.com/ToVoice/3000-33660_4-75901587.html |
| 48. Software | TTSUU | http://download.cnet.com/TTSUU/3000-33660_4-75563194.html |
| 49. Software | TextAloud | http://nextup.com/download.html |
| 50. Software | Nextup Talker | http://nextup.com/download.html |