



Effects of Educational Context on Learners' Ratings of a Synthetic Voice

Neasa Ní Chiaráin, Ailbhe Ní Chasaide

The Phonetics and Speech Lab., School of Linguistic, Speech and Communication Sciences,
Trinity College Dublin, Ireland

neasa.nichiarain@tcd.ie, anichsid@tcd.ie

Abstract

Studies suggest that the context in which synthetic voices are evaluated has a significant bearing on the evaluators' assessments of the voices. The present study is based on learners' evaluations of an Irish synthetic voice which is used in a Computer-Assisted Language Learning (CALL) context.

Previous studies have allowed a number of extraneous variables, such as different voices being used or pitch or speech rate manipulations, to enter the study. The present study involves a single synthetic Irish HTS voice being evaluated in three different contexts, one of which is highly interactive, a second less so and a third which does not require the evaluator to perform any specific task. Results show that the voice was rated more favourably when presented in a highly interactive context. This finding has significant implications for the way we interpret the results of evaluations of synthetic voices.

Index Terms: text-to-speech synthesis, evaluation, computer-assisted language learning, Irish (Gaelic)

1. Introduction

The question of how to evaluate synthetic voices has been a subject of discussion for some time now [1]. Criteria such as intelligibility, naturalness and attractiveness are frequently applied though it is not always clear that those who are assessing the voices have a common understanding of what exactly is intended by each of the terms. Handley [2] suggested that one's judgment of the voices was dependent on the role which those voices had in the context in which they were being judged. Ní Chiaráin [3] suggested that one's evaluation of the voice was dependent on the degree to which one engaged with the situation in which the voice was being used. The latter study involved three distinct environments which were used as CALL platforms, or serious games, for the teaching of the Irish language to sixteen year-old students. One involved navigating a virtual world in order to complete a task, a second had a rich visual environment with a low level of interactivity, while the third involved a dialogue partner, or chatbot, which had relatively simple graphics but was highly interactive. The synthetic voices in question were HTS voices which were developed as part of the ABAIR initiative at the Phonetics and Speech Lab., CLCS, TCD (www.abair.ie).

It emerged that the platform which involved the highest level of interactivity was scored the highest by learners (N=251). Although the same voices were used for each of the platforms, it was necessary to carry out pitch and speed manipulations on them so that different characters in the various platforms could be distinguished from each other.

The present study examines the relationship between learners' evaluations of a synthetic Irish voice and the degree of interactivity and engagement which the learners have with the CALL platform in which the voice is being used. It uses a single voice which has not been manipulated thereby reducing the chance of extraneous variables such as pitch or speech rate changes influencing the outcome. The single voice is evaluated in terms of intelligibility, quality and attractiveness in three diverse CALL contexts:

- (1) a highly interactive context where the learner interacts directly with a chatbot
- (2) a task-based activity where learners are asked to transcribe sentences as they hear them
- (3) a non-task-based activity where learners merely listen and follow the text on the screen

Participants are asked to evaluate the voice after each context. The hypothesis is that ratings for the interactive context (1) will be the highest, followed by the task-based context (2), and that the least highly rated will be the non-task-based activity (3), where the activity is such that learners are least likely to be engaged.

This study has significant implications for the way we interpret evaluations of synthetic speech and may account for inconsistent evaluations of the same voice.

2. Motivation and Rationale for this Study

There is a threefold motivation for the present study. Firstly, it wishes to extend the debate on the evaluation of synthetic speech with particular reference to the connectivity between evaluators' perception of the synthetic voices and the context in which the voices are being experienced.

The second motivation for conducting the study relates to the usefulness of synthetic speech as an aid to language teaching/learning in a CALL context. To date, synthetic speech has been used in a very restricted way for language teaching. Continuous improvements in the quality of the voices could be expected to add greatly to their usefulness as language teaching aids. In the context of interactive CALL platforms, where natural language processing/artificial intelligence is used to create chatbot-type platforms where the computer interacts in a non-linear fashion with the input of human interlocutors, then the synthetic speech is the only type of speech which can be used. The computer can generate an infinite number of responses which could clearly not be pre-recorded. The present study is part of a larger study that aims to evaluate the usefulness of synthetic speech for such a context.

Finally, the study is part of the overall ABAIR initiative, which has developed synthetic voices for the Irish dialects. The Irish language is designated as one of the endangered

languages by UNESCO [4]. There are relatively few native speakers of Irish available as speaker models for learners and the availability of interactive software programs is similarly restricted. The ABAIR initiative aims to develop interactive Irish language teaching/learning platforms into which the synthetic speech is integrated. This study is important to provide an evaluation of the suitability of the synthetic voices for such platforms at this stage in their development.

3. Literature Review

To date, text-to-speech (TTS) synthesis has not been used very frequently as an integral part of Computer-Assisted Language Learning (CALL) systems [5]–[7]. In earlier years this was largely because of the relatively poor quality of the synthetic voices and their inability to closely imitate the human voice [8]. More recent advances in the development of TTS have now reached the point where most systems are almost 100% intelligible [1]. Indeed, in one case, language learners found it difficult to distinguish between natural and synthetic speech [1], [9]. Despite the advances and the vastly improved quality of many of today's synthetic voices the observations of Keller and Zellner-Keller [10] that they are insufficiently expressive and lack any form of human emotion, are still valid to-day.

The greatest weakness being reported in TTS being used for CALL purposes is at the suprasegmental level of production where it lacks appropriate prosody with the result that listeners can experience difficulty in focusing on the output leading to lower levels of comprehension than would be the case for a natural voice [11].

To date, synthetic speech in CALL platforms have tended to be used for the teaching of individual skills such as listening comprehension, reading or pronunciation skills [6], [12], [13]. Modern approaches to language teaching favour the learner being involved in a task in a communicatively interactive way, using the target language in order to achieve some pre-defined goal. Some language learning 'games' are emerging which allow the learner to interact with a virtual interactive dialogue partner, or chatbot, in a non-linear fashion so that so that the output responds to the input of the learner in an individualized fashion. This type of platform is frequently referred to as a 'serious game'. Until recently there have been few instances of the integration of TTS into serious CALL games [14]–[17]. The use of TTS synthetic voices becomes essential in CALL platforms which are genuinely interactive in a non-linear fashion. In this situation pre-recorded voices cannot be used as there may be a near infinite number of possible responses. Since we are at the very initial stages of integrating synthetic voices into interactive CALL platforms with dialogue games we must now revisit the criteria by which TTS is evaluated in order to establish its fitness-for-purpose in this evolving context [18].

King [1] suggests that subjective listening tests are the only sure way to evaluate synthetic speech. Diagnostic or comparative laboratory tests may not equate to the perceptions of the human brain. Furthermore, when the objective is a learning context, it is vital that the adequacy of the voice be assessed by learners, where additional factors such as cognitive load are realistically mimicked. When one introduces subjectivity to any evaluation then context and evaluators' characteristics come into play. The effects of context in visual perception have been well documented and there are numerous examples of how the perception of objects changes when its context changes without any physical

changes in the object itself [19]. This same phenomenon, though less widely reported, occurs in the case of auditory perception. Situational context has been described in terms of the psychological, social, physical, physiological and linguistic constraints which come to bear on the individual's perception of auditory cues [20] [21] [22] [23], [24]. The phenomenon known as the 'McGurk Effect', for example, illustrates the multimodal nature of speech perception since it takes auditory and visual cues in processing an auditory stimulus [25]. Similarly, Munhall et al. noted the relationship between physical head movements and linguistic meaning amongst Japanese speakers [26].

Context and perception are inextricably linked and consequently evaluations of phenomena such as synthetic speech remain somewhat elusive [1]. Evaluators' judgments vary with context, as in the case in the perception of natural language, and the literature draws attention to a variety of factors found to influence perceptions of synthetic speech. It has also been suggested that a learner's evaluation may be quite different to that of a native speaker. Kang et al. [7] found that learners' general listening comprehension abilities influenced their ratings of synthetic voices. Stern et al. [27] suggested that synthetic voices are judged more positively when they are presented as emanating from a computer rather than being human voices. Cryer and Home suggested that listeners' predisposition to synthetic speech based on their previous experiences with it may influence their attitude or prejudices towards it [28]. One may reasonably claim that the context in which synthetic speech is being judged has an influence on the listeners' evaluations of it. Ní Chiaráin suggested that in the case of a CALL platform, one's evaluation of synthetic speech may be highly influenced by one's perceptions of other facets of the platform such as graphics, plot, attractiveness, playfulness, etc. [3].

4. Research Design & Experimental Setup

The present study examines the degree to which learners' perceptions of one Irish TTS synthetic voice (see Section 4.2) vary within three diverse CALL contexts (see Section 4.3).

4.1. Experimental setup

Three schools were selected for participation in this study. These are Irish-medium schools and were selected to ensure participants would have a high level of Irish. This was done to ensure the scores for intelligibility weren't affected by lack of knowledge of the language, although it is important to mention that the students may not necessarily be familiar with the dialect spoken by the HTS system.

It was decided to use a 6-point Likert scale to elicit the subjective intelligibility, quality and attractiveness ratings from the participants. This was intended to avoid the tendency of respondents to opt for the mid-point in a 5-point scale. Selections to the level of one decimal point were made possible. It was felt that allowing ratings to a level of one decimal point would allow respondents to give a more fine-grained response rather than relying on whole numbers. The scale was presented in the form of a sliding bar where the default was set at the midpoint of 2.5 and participants were free to drag the slider to the left towards 0 or to the right towards 5. Meaningful statements label the axes (e.g. Intelligibility: 0 "*impossible to understand*" to 5 "*very easy to understand*"). This configuration allows for the calculation of

a mean score for scale items which can then be used to produce the parametric mean and standard deviation.

It was decided to host the evaluations online so that they could be delivered to students in their own school environment. Each student worked individually and was equipped with a set of headphones to listen to the speech output.

On the evaluation day, students were given a broad general introduction to the area of speech and language technology by the researcher as well as a short introduction to the Irish ABAIR initiative. This entailed explaining that the project aimed to develop multi-dialect synthetic voices for Irish. It was not specified to the students how many voices were involved in the initiative, rather that all three of the main dialects of Irish were represented. They were told that they were evaluating the synthetic speech and that their opinions were valued because of their advanced level of Irish. They were given a half an hour to work individually at their computers and were presented with three diverse CALL contexts in random order. Each context used an embedded synthetic voice. A pre-evaluation questionnaire was presented to students to elicit some basic background information, such as gender and age. Empty text boxes inviting feedback were included on each page of the evaluation.

4.2. One HTS voice

A synthetic voice was built with HTS [29] using a corpus recorded from a middle-aged male speaker of the Connaught dialect of Irish. This voice was built based on c. 3,000 utterances at a 32 kHz sample rate. An example of the synthetic voice can be heard by visiting www.abair.ie, selecting ‘Conamara HTS’ and synthesizing some Irish text.

4.3. Three CALL contexts

The three diverse CALL contexts were presented to students in random order to limit as much as possible fatigue impacting on the scoring of any individual context.

4.3.1. A highly interactive context

Students were given the opportunity to have a playful, non-directed interaction with a chatbot call *Taidhgin* which was developed as a prototype spoken dialogue system by Ní Chiaráin [3]. *Taidhgin* is a chatbot in the form of a monkey who appears as though he is engaging in conversation by typing prompts or responses into a laptop (see Figure 1). Learners make their input by typing into a text box. Irish does not yet have a speech recognition system and hence the only input possible for now is in text form. The chatbot is built on the basis of pattern matching and responds to input from a database of preprogrammed content, which is appropriate to the language level of a 16-17year old learner in the Irish school system. For more details on *Taidhgin* see Ní Chiaráin and Ní Chasaide [18]. Students were not given a specific time limit for interacting with the chatbot but most spent 10-15 minutes in the engagement.

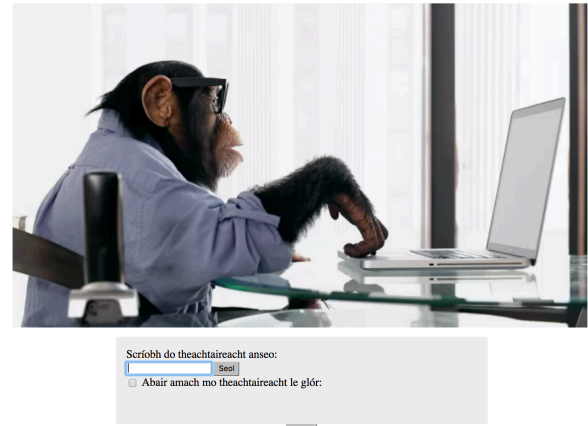


Figure 1: Context 1–*Taidhgin*, the interactive dialogue partner.

4.3.2. A task-based activity

In this activity the evaluators were asked to transcribe 20 sentences as they heard them. The sentences were composed especially for this evaluation – the language used was at an appropriate level of difficulty for the target group, using vocabulary which belongs in the domain of schooling and is commonplace for them. Once synthesized, there were no modifications made to the output. All sentences were unseen and students were asked to orthographically transcribe them. Each sentence was decontextualized and presented in random order with no common semantic connections between them. Students were free to listen repeatedly, if necessary. The average sentence length was 13.45 words. The time spent was comparable to the interactive context (c. 10-15 minutes).

Ceist 1

0:00 / 0:04

(1) Déan transcriobh ar an méid a chuala tú:

(2) 3 cheist faoin nguth: Tabhair marc idir 0 - 5 d'intuigtheacht (intelligibility), do chaighdeán (quality) agus do tharraingteacht (attractiveness) an ghutha.

i. **Intuigtheacht / Intelligibility** (do dhéanta le tuiscint) 0 - 5 (an-éasca le tuiscint)

ii. **Caighdeán labhartha na Gaeilge / Quality** (an-lag) 0 - 5 (an-ard)

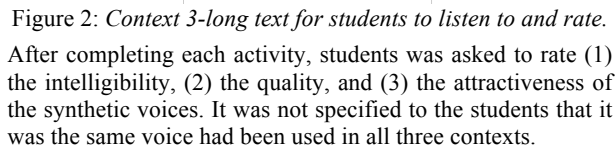
iii. **Tarraingteacht an ghutha / Attractiveness** (gan a bheith tarraingteach) 0 - 5 (guth an-tarraingteach)

An bhfuil aon aiseolas eile a ritheann leat?

Figure 2: Context 2–presentation of the twenty utterances for individual transcription & rating.

4.3.3. A non-task-based activity

Students were asked to listen to a continuous piece of prose enunciated by the synthetic voice. Unlike the sentences in the previous context, these formed a coherent whole and evaluators could also read the script on screen. It was comprised of 421 words divided into 3 separate sections and lasted 2 minutes 23 seconds. As in the previous two contexts, the content and difficulty level were considered appropriate and there was no restriction on the number of repeats they were allowed to make. Typically, this task took about 5 minutes.



2.5) point to a fairly neutral judgment, as being neither particularly attractive/unattractive.

6. Discussion

The hypothesis is supported by these findings: one HTS voice, several CALL contexts, evoked different reactions. This finding highlights the need to reconsider how and why we evaluate synthetic speech and provides further evidence that context influences our perception of auditory cues.

Of the three measures used in this study, intelligibility is undoubtedly the most critical. It is probably also the most objective since subjective evaluations can be supported by more objective indicators, such as the results of transcription tasks, as was done here. It is interesting to note that while some students got a perfect or near-perfect performance score in the transcription task, they nevertheless evaluated the intelligibility and quality of the speech as falling short of perfect. This may be interpreted as indicating that while the speech is intelligible, nevertheless it may take greater mental effort to understand the synthetic voices than would be the case for natural voices. This greater effort gives rise to what is termed “cognitive loading” [30] and can be expected to depress subjective evaluation scores. Ní Chiaráin & Ní Chasaide used a formula to produce an intelligibility and clarity index which averaged the subjective and objective scores and argued that this produced a far more realistic representation of the intelligibility of synthetic speech than either measure on its own [31].

Intelligibility is the single most important factor when one is using synthetic speech for interactive language learning games in that it allows the learner to interact with the chatbot in a coherent manner. While intelligibility cannot be sacrificed, a relatively lower rating for quality may not be as critical as it has been shown that learners are more tolerant of the voice when emanates from a machine [28].

The judgments on attractiveness are the lowest of the three measures. One potentially important factor here is that the voice was of a middle-aged male, of rural background and the evaluators were predominantly female adolescents from an urban background. Clearly, the voice they were rating is unlikely to conform to their image of ‘cool’ or ‘attractive’. (We speculate that very different and more positive responses might be obtained from older evaluators from a rural background). Attractiveness is highly subjective and is likely to vary in a way that reflects the evaluation group. Importantly, in the context of game-characters it is only one of many potential attributes one would wish for. For example, an authoritative, formal, friendly, timid or playful voice may be more appropriate depending on the role which is being assigned to the game character to make it more credible. Given that listeners make subjective judgments about personality based on the quality of the voice [32], [33], the future for CALL may be to match the synthetic voice(s) to the character-role(s). As we move towards more sophisticated interactive dialogue-based CALL scenarios, our evaluations may need to be broadened to explore the perceived age and ‘personality’ of the character conjured by the synthetic voices at our disposal. Looking further into the future, we would also aspire to voices where the quality can credibly be modulated in ways that are appropriate to the context.

On the question of evaluation metrics, it is worth considering whether the widely-used ‘intelligibility’, ‘quality’ and ‘attractiveness’ are the right measures of synthesis quality. Other metrics can also be very important depending on the

circumstances. In the Irish context, for example, ‘nativeness’, i.e. the native-like quality of the voice is a priority for many, given that the interactive CALL applications aim to remedy the lack of availability of native-speaker models of the language. But beyond the realm of the endangered language, nativeness/authenticity can also be important in assistive technologies, where the synthetic voice becomes the user’s voice.

7. Conclusions

The present results support the initial hypothesis that the context in which the voice is used will affect how it is perceived by the user. It seems clear that the highly interactive playful chatbot was more engaging and that this in itself enhanced listeners’ attitudes towards the synthetic voice.

Results also indicate that the synthetic Irish voice that has been developed as part of the ABAIR initiative is fit-for-purpose for interactive CALL activities and has a very high level of intelligibility among 16-17 year old learners of Irish. In the absence of a wide variety of synthetic voices, as in the case of minority languages, one may still proceed with a voice that is highly intelligible so that learners can engage with the target language in an interactive, task-based manner [34].

The voice developed here is but one of three current HTS ABAIR voices, spanning the three major dialects and offering choice in terms of gender and age. As further voices are developed, a priority is to ensure that they measure up to the contexts in which they are to be used and for this early consideration of appropriate evaluation criteria will be critical.

8. Acknowledgements

Funding from the Irish Government’s Department of Arts, Heritage, Regional, Rural and Gaeltacht Affairs (ABAIR Initiative) and An Chomhairle um Oideachas Gaeltachta agus Gaelscolaíochta (CabairE project) is gratefully acknowledged. Buíochas mór leis na múinteoirí agus leis na daltaí scoile a ghlac páirt sa taighde seo.

9. References

- [1] S. King, “Measuring a decade of progress in Text-to-Speech,” *Loquens*, vol. 1, no. 1, pp. 2386–2637, 2014.
- [2] Z. Handley, “Evaluating text-to-speech (TTS) synthesis for use in computer-assisted language learning (CALL) (Unpublished doctoral dissertation),” University of Manchester, UK, 2005.
- [3] N. Ní Chiaráin, “Text-to-Speech Synthesis in Computer-Assisted Language Learning for Irish: Development and Evaluation,” (Doctoral thesis, CLCS, Trinity College, Dublin), 2014.
- [4] C. Moseley, Ed., *Atlas of the world’s languages in danger*, 3rd ed. Paris: UNESCO Publishing, 2010.
- [5] P. Gupta and M. Schulze, “Human language technologies (HLT),” *Human Language Technologies (HLT). Module 3.5 in Davies, G. (Ed.) Information and Communications Technology for Language Teachers (ICT4LT)*, Slough, Thames Valley University [Online], 2012. [Online]. Available: http://www.ict4lt.org/en/en_mod3-5.htm. [Accessed: 30-Oct-2013].
- [6] B. Guclu and S. Yigit, “Using Text to Speech Software in Teaching Turkish for Foreigners : The Effects of Text to Speech Software on Reading and Comprehension Abilities of African Students,” *J. Humanit. IBSU*, vol. 4, no. 2, pp. 31–33, 2015.
- [7] M. Kang, H. Kashiwagi, J. Treviranus, and M. Kaburagi, “Synthetic speech in foreign language learning: an

- evaluation by learners,” *Int. J. Speech Technol.*, vol. 11, pp. 97–106, 2008.
- [8] J. Higgins, *Language, learners, and computers: Human intelligence and artificial unintelligence*. New York: Longman, 1988.
- [9] T. Bione, J. Grimshaw, and W. Cardoso, “An evaluation of text-to-speech synthesizers in the foreign language classroom: learners’ perceptions,” in *CALL communities and culture – short papers from EUROCALL 2016*, S. Papadima-Sophocleous, L. Bradley, and S. Thoušny, Eds. Research-publishing.net, 2016, pp. 50–54.
- [10] E. Keller and B. Zellner-Keller, “Speech synthesis in language learning: challenges and opportunities,” in *Proceedings of InSTIL 2000 Conference*, 2000, pp. 109–116.
- [11] M. Wester, O. Watts, and G. E. Henter, “Evaluating comprehension of natural and synthetic conversational speech,” in *Proceedings of the 8th International Conference on Speech Prosody (SP2016)*, 2016, no. August, pp. 766–770.
- [12] D. González, “Text-to-speech applications used in EFL contexts to enhance pronunciation,” *TESL-EJ*, vol. 11, no. 2, pp. 1–11, 2007.
- [13] F. Kiliçkaya, “Improving pronunciation via accent reduction and text-to-speech software,” in *WorldCall: International Perspectives on Computer-Assisted Language Learning*, M. Levy, F. Blin, C. Bradin Siskin, and O. Takeuchi, Eds. New York: Routledge, 2011, pp. 85–96.
- [14] N. Peirce and V. Wade, “Personalised learning for casual games: The ‘Language Trap’ online language learning,” in *Proceedings of the Fourth European Conference on Game Based Learning (ECGBL 2010)*, 2010, pp. 306–315.
- [15] N. Ní Chiaráin and A. Ní Chasaide, “The Digichaint interactive game as a virtual learning environment for Irish,” in *CALL communities and culture – short papers from EUROCALL 2016*, S. Papadima-Sophocleous, L. Bradley, and S. Thoušny, Eds. Research-publishing.net, 2016, pp. 330–336.
- [16] P. Su, Y. Wang, T. Yu, and L. Lee, “A Dialogue Game Framework with Personalized Training Using Reinforcement Learning for Computer-Assisted Language Learning,” in *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*, 2013, pp. 8213–8217.
- [17] H. Morton, N. Gunson, and M. Jack, “Interactive language learning through speech-enabled virtual scenarios,” *Adv. Human-Computer Interact.*, vol. 2012, no. Article ID 389523, p. 14 pages, 2012.
- [18] N. Ní Chiaráin and A. Ní Chasaide, “Chatbot Technology with Synthetic Voices in the Acquisition of an Endangered Language: Motivation, Development and Evaluation of a Platform for Irish,” in *the Tenth International Conference on Language Resources and Evaluation (LREC 2016)*, 2016.
- [19] D. Todorović, “Context effects in visual perception and their explanations,” *Rev. Psychol.*, vol. 17, no. 1, pp. 17–32, 2010.
- [20] J. L. McClelland and J. L. Elman, “The TRACE model of speech perception,” *Cogn. Psychol.*, vol. 18, pp. 1–86, 1986.
- [21] J. Hay, P. Warren, and K. Drager, “Factors influencing speech perception in the context of a merger-in-progress,” *J. Phon.*, vol. 34, no. 4, pp. 458–484, 2006.
- [22] K. M. Hutchinson, “Influence of sentence context on speech perception in young and older adults,” *J. Gerontology Psychol. Sci.*, vol. 44, no. 2, pp. 36–44, 1989.
- [23] F. Grosjean and J. P. Gee, “Prosodic structure and spoken word recognition,” *Cognition*, vol. 25, no. 1–2, pp. 135–155, 1987.
- [24] B. Spehar, S. Goebel, and N. Tye-Murray, “Effects of Context Type on Lipreading and Listening Performance and Implications for Sentence Processing,” *J. Speech, Lang. Hear. Res.*, vol. 58, no. 3, pp. 1093–1102, 2015.
- [25] H. McGurk and J. MacDonald, “Hearing lips and seeing voices,” *Nature*, vol. Dec 23-30; no. 264(5588), pp. 746–8, 1976.
- [26] K. G. Munhall, J. a Jones, D. E. Callan, T. Kuratate, and E. Vatikiotis-Bateson, “Visual prosody and speech intelligibility: head movement improves auditory speech perception,” *Psychol. Sci. a J. Am. Psychol. Soc. / APS*, vol. 15, no. 2, pp. 133–137, 2004.
- [27] S. E. Stern, J. W. Mullennix, and I. Yaroslavsky, “Persuasion and social perception of human vs. synthetic voice across person as source and computer as source conditions,” *Int. J. Hum. Comput. Stud.*, vol. 64, pp. 43–52, 2006.
- [28] H. Cryer and S. Home, “User attitudes towards synthetic speech for talking books,” RNIB Centre for Accessible Information, Birmingham: Research report #7., 2009.
- [29] K. Tokuda, H. Zen, and A. W. Black, “An HMM-based speech synthesis system applied to English,” in *IEEE Speech Synthesis Workshop*, 2002, pp. 227–230.
- [30] H. K. Pillay, “Cognitive load and mental rotation: structuring orthographic projection for learning and problem solving,” *Instr. Sci.*, vol. 22, pp. 91–113, 1994.
- [31] N. Ní Chiaráin and A. Ní Chasaide, “Evaluating text-to-speech synthesis for CALL platforms,” in *Antwerp CALL 2014: International CALL Research Conference*, 2014, pp. 104–110.
- [32] P. McAleer, A. Todorov, and P. Belin, “How do you say ‘hello’? Personality impressions from brief novel voices,” *PLoS One*, vol. 9, no. 3, pp. 1–10, 2014.
- [33] K. R. Scherer, “Personality inference from voice quality: The loud voice of extroversion,” *European Journal of Social Psychology*, vol. 8, no. 4, pp. 467–487, 1978.
- [34] R. Ellis, *Task-based language teaching and learning*. Oxford: Oxford University Press., 2003.