Mobilizing Instruction in a Second-Language Context: Learners' Perceptions of Two Speech Technologies

Denis Liakin [1,*], Walcir Cardoso [2] and Natallia Liakina [3]

[1] Département d'études françaises, Concordia University, 1455 De Maisonneuve Blvd. W, Montreal, QC H3G 1M8, Canada

[2] Department of Education, Concordia University, 1455 De Maisonneuve Blvd. W, Montreal, QC H3G 1M8, Canada; walcir.cardoso@concordia.ca

[3] French Language Centre, McGill University, Arts Building, Room 265. 853 Sherbrooke Street West, Montreal, QC H3A 0G5, Canada; natallia.liakina@mcgill.ca

* Correspondence: denis.liakin@concordia.ca; Tel.: +1-514-848-2424 (ext. 7508)

Abstract: We report the results of two empirical studies that investigated the use of mobile text-to-speech synthesizers (TTS) and automatic speech recognition (ASR) as tools to promote the development of pronunciation skills in L2 French. Specifically, the study examined learners' perceptions of the pedagogical use of these tools in learning a French segment (the vowel /y/, as in *tu* 'you') and a suprasegmental feature (across-word resyllabification/liaison, observed in *petit enfant* 'small child'), in a mobile-assisted context. Our results indicate that, when used in a "learn anytime anywhere" mobile setting, the participants believe that they have: (1) increased and enhanced access to input; and (2) multiple opportunities for speech output and (3) for the development of prediction skills. Interestingly, these findings meet the requirements for successful L2 learning, one that recommends the inclusion of pedagogical activities that promote exposure to input (Nation & Newton 2009), multiple opportunities for output (Swain 1995), and the development of prediction skills (Dickerson 2015) to foster learner autonomy and, consequently, to maximize classroom time by extending the reach of the classroom. Our findings also indicate that participants recognize the pedagogical importance of TTS and ASR, and enjoy the mobile-enhanced learning environment afforded by these two technologies.

Keywords: text-to-speech synthesizers; automatic speech recognition; mobile learning; input; output; prediction; French pronunciation; learners' perception

1. Introduction

There seems to be a consensus in current second/foreign language (L2) pronunciation research that one of the ultimate goals of pronunciation instruction is to assist students in improving their intelligibility so that they can understand and be understood by other speakers of the target L2 [1–4]. To achieve these goals, students must engage in a number of learning activities that include at least five phases that go from sound analysis and awareness raising, to listening discrimination, and finally production, as recommended by Celce-Murcia et al. [1] (p. 45): (1) description and analysis of the target feature; (2) listening discrimination; (3) controlled practice; (4) guided practice; and (5) communicative practice. The authors also recommend that the application of these phases be extended over the course of several lessons. The main constraint in following Celce-Murcia et al.'s recommendation to help students attain reasonable levels of intelligibility is time [5]. In a typical foreign learning context, for example, L2 instruction is restricted to one to three hours per week. In these dire circumstances,

how can teachers and students make better use of their classroom time so that they can minimize the time limitations described above?

One way of addressing time-related limitations is by mobilizing the L2 learning experience by promoting "anytime anywhere" learning via mobile devices such as the ubiquitous smartphone, a device that is easily available to most students [6]. According to the Pearson Student Mobile Device Survey conducted in 2015 [7], for instance, 86% of university students in the United States use smartphones frequently. In addition to the widespread use of smartphones, they have been praised for their pedagogical potential, according to a study published by UNESCO [8]; this includes the ability to personalize learning, to extend educational practices and experiences beyond classrooms, and to allow for both informal and teacher-assisted (hybrid) learning. The ongoing expansion of smartphone use has sparked increasing interest by CALL/MALL (computer-assisted language learning/mobile-assisted language learning) researchers and practitioners, with most studies emphasizing their potential for a mobile-assisted L2 pedagogy [6,9–15].

In the context of L2 pronunciation, and assuming that learning involves both exposure to the target language (i.e., input; phases 1 and 2 in Celce-Murcia et al.'s framework for teaching pronunciation, described above [1]) and oral practice (i.e., output; phases 3, 4 and 5 in the same framework), two speech technologies have the potential to promote both input and output practice, in a mobile platform: Text-to-Speech Synthesizers (TTS) and Automatic Speech Recognition (ASR). Because these technologies make use of written text (orthography), they also play an important role in the development of spelling-to-sound prediction skills, as advocated by Dickerson [16,17]. In sum, the combined use of TTS and ASR has the potential to enhance the language learning experience by providing learners with increased and enhanced input, many opportunities for output practice, and the development of spelling-to-sound predictions.

While research on the pedagogical uses of TTS and ASR in L2 education is still in its infancy, the studies currently available indicate positive effects after their prolonged use, particularly as complements of classroom instruction. For example, Liakin et al. [10,11] have shown that by mobilizing the learning of L2 French with TTS and ASR using smartphones (i.e., for the acquisition of the vowel /y/ as in *tu* 'you' and the liaison observed in *peti/t.e/nfant > peti/.te./nfant* 'small child', where the period indicates syllable boundaries), learners were able to significantly improve their performance in the two phenomena investigated, in comparison with the respective control groups (see details in forthcoming discussion). Although these results were optimistic from a learning perspective, one question remains to be answered: are L2 students willing to mobilize their learning experience by using TTS and ASR as pedagogical tools to learn about French pronunciation? What are their perceptions of and attitudes towards such an approach to learning, one that emphasizes the mobilized use of TTS and ASR? These are the questions that we will address in this study, which is part of a larger investigation of the pedagogical use of speech technologies such as TTS and ASR in L2 education.

2. Background

As indicated earlier, one of the problems that afflict the L2 classroom is the amount of time that can be dedicated to teaching specific items of the target language, particularly pronunciation features. One way to circumvent these time limitations is to provide learners with tools that allow them to learn on their own, beyond the walls of the language classroom, and that encourage them to focus on activities that motivate learning. These include access to the target language's input and providing opportunities for production (output) and the development of prediction skills. In this section, we introduce these important goals of pronunciation instruction, and discuss research on two speech technologies that have the potential to mitigate the time limitations described and, consequently, promote L2 pronunciation learning: TTS and ASR. Accordingly, we contextualize and motivate the scope and goals of our study: to examine learners' perceptions of the use of TTS and ASR as self-learning tools for the development of L2 French pronunciation skills.

2.1. Input, Output, Prediction and Time Constraints

Research on pronunciation instruction suggests that L2 instructions should target the development of learners' linguistic competence in the target L2 system. According to Dickerson [16,17], for example, this competence includes learners' ability to perceive (input), produce (output), and predict pronunciation features (based on orthographic input-output relationships). These three competence elements constitute what Dickerson defines as the trilogy of goals (or 3Ps): perception (input), production (output) and prediction.

The input notion is a construct directly associated with perception because, without access to items in the target language (input), learners cannot develop their perceptual skills. As such, it plays a pivotal role in acquisition, particularly if it is comprehensible [18,19]. While its importance has been advocated by many linguistic theories (e.g., it serves to set parameters in generative approaches to linguistics [20]), this hypothesis is based on the assumption that learners' ability to recognize or perceive features in the target language is developed before their ability to produce them (e.g., [21,22]—see also [1] for a similar view in the context of pronunciation teaching).

While the input is important in the development of L2 pronunciation, as discussed above, production (or output) also plays a vital role in language development. According to Swain [23], for example, output practice has the ability to aid learners in the development of automacity, noticing abilities, and hypothesis testing, wherein learners use language production to test their assumptions about the language being learned so that they can understand how it functions. Without opportunities for output, L2 learners cannot practice language, nor can they perfect or automatize their linguistic skills [23–25].

Finally, the third item in Dickerson's trilogy of goals [16,17] for the development of linguistic competence in L2 pronunciation instruction is prediction. This goal or approach to pronunciation teaching assumes that orthography is one of the only resources available to L2 learners, since writing is often part of the language curriculum; this allows students to apply what they already know without the burden of learning a phonetic alphabet or other symbols. In addition, orthographic rules have been part of pronunciation textbooks for at least seven decades and, more importantly, spelling-to-sound rules apply systematically in English (e.g., *c* is consistently pronounced as [s] when followed by the letter *e* and *i*—see [1], for a similar conclusion). The main criticism against orthography-based approaches to pronunciation instruction comes from the widespread perception that English orthography is "a chaotic concoction of oddities without order and cohesion" [26] (p. 1). While a similar perception holds for French, the target language in our study ("there is no such systematic correspondence between orthography and pronunciation in French" [27] (p. 3)), we are confident that the "chaotic concoction of oddities without order and cohesion" applies to a much lesser degree in the language of Molière.

Arguably, the most important constraint in fulfilling Dickerson's trilogy of goals for pronunciation instruction is time, which is often limited in the language classroom (from 1–2 h a week in primary and secondary schools in Canada) and consequently insufficient for helping students to attain reasonable levels of L2 proficiency [5]. One way of addressing these limitations is via the use of speech technologies, preferably in mobile formats because they can promote "anytime anywhere learning". In our study, to address the three Ps of Dickerson's trilogy of goals, we adopt the following two speech technologies: TTS (input development) and ASR (output practice) which, with their direct interaction with orthography (e.g., TTS reads written texts, ASR outputs written texts), can motivate the development of prediction skills. As will be clarified in forthcoming sections, orthography is essential for both TTS (to increase and enhance the L2 input) and ASR use (to promote output practice): while TTS increases and enhances the L2 input via spelling-to-sound associations, ASR promotes sound-to-spelling interactions. Combined, these two orthography-based activities may promote the development of prediction skills. A visual representation of our proposal is illustrated in Figure 1, where we show how the adopted speech technologies and orthography work together to accomplish the three goals of L2 pronunciation instruction.

To summarize, along the lines of Dickerson's proposal [16,17], our study incorporates technological tools that address the development of L2 perception, production and prediction via the combined use of two speech technologies, TTS and ASR, which interact with orthography to scaffold the learning process. The effects of these two technologies on L2 learning will be discussed next.
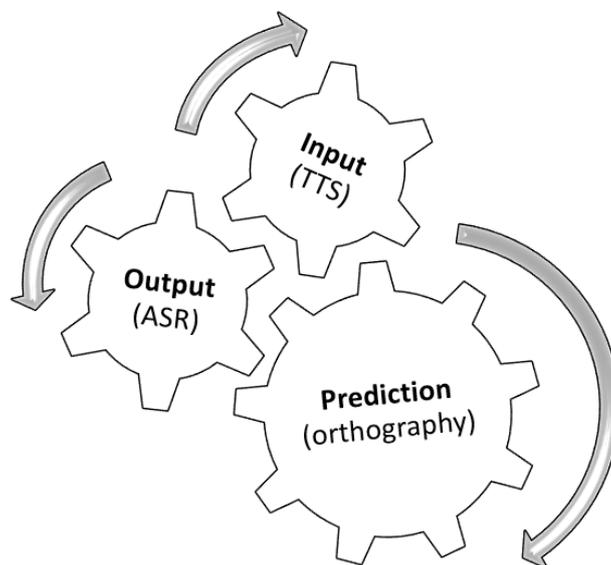

Figure 1. A proposal for the implementation of Dickerson's [16,17] trilogy of goals (3Ps) in a CALL/MALL (computer-assisted language learning/mobile-assisted language learning) context. ASR: automatic speech recognition; TTS: text-to-speech synthesizers.

2.2. Text-to-Speech Synthesizers and Effects on L2 Learning

Text-to-speech synthesizers transforms units of text into speech for audio presentation (e.g., via a pair of speakers or headphones). By rendering audio output from digital text, TTS is commonly used to assist users who cannot read or see, or to provide answers to voice-activated searches such as those found in browsers and GPS systems. Most TTS programs include features that allow for different speech output speeds, gendered voices, contrasting pitches (high and low), various accents, and a function that highlights the word or section of text being read by the program.

There exists very little research on the effects of the use of TTS as a L2 pedagogical tool. Despite this scarcity, the ones currently available do indicate that TTS could be valuable in teaching L2 pronunciation [28–30], specifically in raising learners' awareness to certain language features. Involving other skills, studies have shown that TTS has a positive impact on the learning of writing [31], vocabulary, and reading [32,33]. Notably, we are only aware of one involving its use in *mobile* devices such as smartphones, as will be discussed later.

One question that has not received careful attention in the CALL/MALL literature is whether the quality of TTS synthesis is appropriate for the L2 classroom. This is exactly what Cardoso et al. [29] aimed to investigate in the context of learning L2 English pronunciation. Their findings confirmed that the quality of a commercial TTS voice resembles considerably that of a human in a number of features: learners' ability to comprehend a story, to listen and repeat a short oral text using "shadowing" techniques (in which participants repeat speech immediately after hearing it), and to notice and identify specific phonological features in the target L2 (i.e., regular past tense marking and its allophones as in *added*, *talked* and *played*). Given TTS's remarkable improvement in quality over the years [34], now is the time to take advantage of synthesized speech's potential as an effective pronunciation model for L2 instruction, specifically in a mobile "anytime anywhere" learning environment.

2.3. Automatic Speech Recognition and Effects on L2 Learning

Automatic Speech Recognition (also known as "speech to text" or "computer speech recognition") is a computing process that instantly transcribes spoken language into text. In the context of pronunciation instruction, researchers propose using ASR to teach the pronunciation of a foreign language and to assess students' oral production. While many studies show that ASR technology can be effective for the teaching of segments [35–41], there is a lack of research reporting on learners' perceptions of using such tools.

Both speech technologies (TTS and ASR) have the potential to enhance and contribute to learning. However, if used in static (non-mobile) environments, they limit their use to the lab or to personal computers in one's home or office. One way of motivating ubiquitous learning and to accomplish the goals of our study is via mobile TTS and ASR, as will be discussed next.

2.4. Mobile Speech Technologies and L2 Learning: Automatic Speech Recognition and
Text-to-Speech Synthesizers

Due to their pedagogical potential for use inside and outside the language classroom, mobile devices such as smartphones and iPads (Apple Inc., Cupertino, CA, USA) have caught the interest of a growing number of researchers over the last decade [9–11,13,14,42]. In addition to being useful in language learning, mobile devices can use their multimedia capabilities to create authentic learning experiences for students, thus situating learning within appropriate cultural and linguistic schemata [43].

To our knowledge, there are only two empirical studies that investigated the use of *mobile* TTS and ASR as tools to promote the development of pronunciation in L2 French and the effects of these two types of speech technologies on the learning of L2 phonological features. While one investigated the phenomenon of French liaison [11], the other focused on the acquisition of the French vowel /y/ [10]. The main goals of these studies were to find out whether the use of these technologies, outside of the L2 classroom, contributed to the learning of the two features.

Liakin et al. [11] investigated the acquisition in production of French liaison (the pronunciation of a latent word-final consonant across two words, thus triggering a resyllabification process with the following vowel-initial word (*sans ami* 'without a friend' is pronounced s⁀_[za]mi and not s[⁀_a]mi) in a mobile TTS-based learning environment (using NaturalReader (NaturalSoft Ltd., Richmond, BC, Canada)). Using a pre-/post-/delayed post-test design with two experimental groups (one taught by TTS and another by a French instructor) and a control group, the results indicated that the two groups that received instruction, namely the TTS and teacher-led groups, outperformed the control group in liaison production, thus confirming TTS' ability to aid in pronunciation learning.

Liakin et al. [10], on the other hand, investigated the effects of mobile ASR-based learning (using Nuance Dragon Dictation (Nuance Communications Inc., Burlington, MA, USA)) on the acquisition of the problematic French vowel /y/ (the pronunciation of the letter *u* in words such as *tu* 'you' and *salut* 'hi'), in both production and perception. Using the same research methods utilized in the TTS study described above, the study consisted of three groups of learners: one received instruction via ASR, the other via a French instructor, and the third acted as the control group. Their findings indicated that the group that received ASR-based instruction improved significantly in /y/ production from pre-test to post-test, in comparison with the two other groups. As such, these results support the hypothesis that the pedagogical use of a mobile version of ASR may have a positive effect on /y/ production. As for the perception of /y/, the results showed that despite slightly greater gains for the ASR group, the three groups behaved in a similar way. This indicates that the group that received ASR-based treatment was not able to extend the newly acquired knowledge detected in production to perception.

The results of these two studies show that TTS and ASR, when used in an "anytime anywhere" mobile learning setting, promote the acquisition of French pronunciation, at least in production. But are L2 learners ready and willing to use these technologies to enhance their own learning? The main goal of this study is to answer this question by investigating learners' perception of these two

technologies, used as pedagogical tools to enhance their learning experience beyond the walls of their language classrooms.

The following question and its sub-component guided this research:

1.　What are learners' perceptions of the use of mobile TTS and ASR as pronunciation learning tools (based on a set of 10 pre-determined criteria, to be described later)?
2.　What are the perceived strengths and weaknesses of using TTS/ASR for learning pronunciation in a second language environment, particularly with respects to the tool's ability to provide access to input, increase output opportunities, and develop learners' prediction skills?

To answer these two interrelated questions, we adopted a combination of quantitative measures (i.e., survey results, using descriptive statistics) and qualitative analyses (i.e., open-ended oral interviews with the participants, using Saldaña's and L'Écuyer's protocols [44,45] for analyzing interviews), as will be described in the next section.

3. Method

3.1. Participants

Sixty-nine students of French as a second language participated in this study (27 from the TTS study and 42 from the ASR study—see forthcoming discussion). All participants were recruited from intact L2 French classrooms at two Anglophone universities in Montreal. They were either native English speakers or had native-like proficiency in the language. In that case, the participants' first languages included Mandarin, Arabic and Spanish. To participate in the study, they were required to have elementary level proficiency in French (A2 level, according to the Common European Framework of Reference for Languages). Among the participants who engaged in the pedagogical use of one of the two technologies ($n = 23$), 17 were female and six were male, with an age range of 19–25. As will be clarified later, note that the number of students who participated in the survey and interviews ($n = 23$) is considerably lower than the total number of participants ($N = 69$) because of the nature of this large-scale study, which consisted of an alternative experimental group (teacher-led) and a control group. Evidently, only the 23 participants who used one of the target technologies were asked to engage in further investigation on their perceptions of using TTS and ASR to learn features of French phonology.

As is typical in research of this nature, standard research ethics protocols were observed; for instance, all participants had to read and sign a consent form agreeing to participate in the study. The consent form included items that gave the participants the option to decline participation at any stage of the study, without negative consequences; they were also guaranteed that their participation was confidential and only the researchers and the assistants would have access to the interview materials.

3.2. Design of the Study, Experimental Groups and Treatment

Following Chapelle's recommendation for conducting methodologically convincing CALL research [46,47], this study followed a mixed-methods approach, using a pre/post/delayed post-test research design (quantitative) followed by surveys and interviews with the participants (qualitative). Due to the scope of this study and its main goals, the focus will be primarily on the analysis of the survey results (quantitative) and interview responses (qualitative). The study took place within a one-year period, with similar research methods employed in the two (TTS and ASR) groups.

The participants in the TTS study were randomly assigned to one of three distinct groups, each corresponding to one of the experimental groups adopted: a group that used TTS as a means of instruction (TTS; $n = 9$), a group taught by a French teacher (NTTS; $n = 9$), and a control group (CTL; $n = 9$). The same rationale was used to assign participants in the ASR study: they were randomly assigned to one of three distinct groups, each corresponding to an experimental group: ASR ($n = 14$), NASR ($n = 14$, in which the app was replaced by a teacher) and CTL (control; $n = 14$).

During the treatment period, the participants were not informed about the nature of the studies, except that it was about "an app that could help second language learners improve their French". Figure 2 illustrates the general design of this project (combining both studies), which will be discussed in detail below.

Students in the TTS/ASR Groups practiced French pronunciation using mobile TTS/ASR apps on a tablet or smartphone using two free commercial and easily available applications: NaturalReader (for the TTS study) and Nuance Dragon Dictation (for the ASR study).
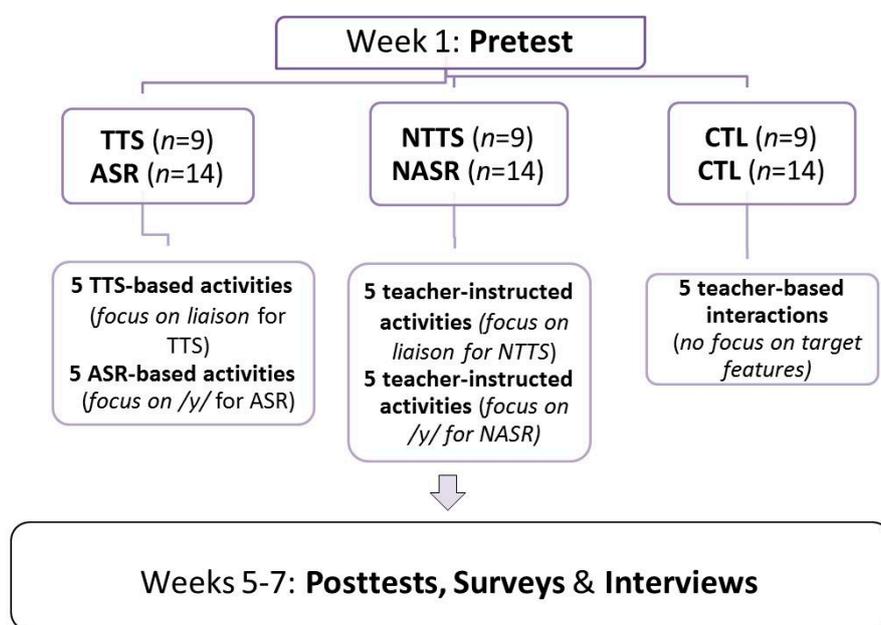

Figure 2. Design of the study. TTS = Text-to-Speech Group; NTTS = Non-TTS Group (teacher-led); ASR = Automatic Speech Recognition group; NASR = Non-ASR group (teacher-led); CTL = Control Group.

The TTS Group participants were asked to practice for approximately 20 min each week, either at home or at the university. Every Monday, the participants were emailed a web link containing the list of weekly expressions that they had to copy and paste into the TTS app. To ensure that the participants completed the activities, they were asked to fill in weekly reports. In this report, the participants were asked to (1) rank how easy/hard it was to understand what they heard; (2) find out which words or phrases contained certain French phonetic features, segments, or phonological phenomena (e.g., whether certain two-word combinations were pronounced together, whether a given word contained a nasal vowel); (3) pick the best-sounding word, the worst-sounding word, and the most difficult word to pronounce.

The ASR Group participants also completed, on a weekly basis, five 20-min pronunciation activities that consisted of reading aloud the target words and phrases in French using the ASR software installed on their mobile devices. After each reading attempt, students were provided with immediate written visual feedback via an orthographic representation of their attempt. To illustrate, if students attempted to pronounce the word *pure* 'pure' [pyr] but *pour* 'for' or *pire* 'worst' appeared on their screen as the written result (i.e., the ASR visual output), this indicated that their pronunciation was incorrect, thus requiring another attempt. The ASR participants were asked to spend approximately one minute per word/phrase, depending on the level of difficulty of each target phrase, for a total of 20 min. The rationales behind our decision to encourage learners to spend approximately one minute per word or phrase are two-fold: (1) in our pre-experiment observations (piloting stage), this is the amount of time we found, on average, appropriate for each question; and (2) it ensures that all groups receive the exact same amount of time for practice. To ensure that the participants completed the

assigned ASR-based pronunciation activities, they were also asked to indicate, on a "pronunciation form" (weekly report) the number of times they repeated each form until they were able to produce it accurately, or until their 1-min limit had expired.

To better contextualize our research and for the sake of completion, we provide now a description of the treatments that the teacher-led and control groups received in each study. Contrary to what was observed in the TTS and ASR studies, participants in the teacher-led NTTS and NASR Groups did not have access to mobile devices. However, they completed the same activities as the TTS and ASR participants in weekly 20-min sessions with a French instructor who: read the sentences/phrases to the participants, asked them to repeat each sentence/phrase at least three times (e.g., so that they could decide, for example, which words or phrases contained certain French features, including segments, phonetic and/or phonological phenomena), and asked them to select the best-sounding word, the worst-sounding word, or the most difficult word to pronounce. To accommodate the nature of the intervention received by this group, the weekly reports were slightly adapted (e.g., the irrelevant parts like "try to understand what the app says" were removed).

Finally, the Control Group participated in weekly individual 20-min meetings with a French instructor, wherein the participant and the teacher engaged in discussions on a variety of topics about school, aspirations, family, etc. with the goal of practicing their conversation skills. The teacher provided no feedback on liaison (TTS study) or /y/ pronunciation (ASR study) or any other pronunciation features.

3.3. Materials

A questionnaire survey (adapted from [48]) and an open-ended oral interview were adopted to collect data on students' perceptions of the use of the TTS and ASR software. The questionnaire consisted of multiple-choice items using a five-point Likert scale in order to measure the degree to which students disagreed or agreed with each statement: (1) strongly disagree; (2) disagree; (3) neutral; (4) agree; and (5) strongly agree. There were 10 items targeting specific questions about the students' perceptions of and attitudes towards the use of the TTS and ASR software in this project. These 10 criteria were selected because they have proved to capture a number of issues that may affect L2 learners' perceptions of technology (see [48] for details) and, more importantly, they are in line with the goals and scope of this research.

To guarantee confidentiality and to avoid factors that could affect data collection or interpretation of the statements (e.g., the presence of the teacher, ignorance of a pertinent English word), the survey was administered at home, without the presence of the teacher, and using English, the language of instruction at the university where the study took place.

For the open-ended oral interview, the participants engaged in a conversation in English with one of the researchers about the following questions: "Can you tell us about your experience using TTS/ASR as a learning tool in this project?" and "In your experience, what are the perceived strengths and weaknesses of using the TTS/ASR in this project?" Whenever necessary, the participants' responses were accompanied by related follow-up questions.

3.4. Procedure and Data Analysis

At the end of the post-test (Week 5), participants in the TTS group ($n = 8$) and ASR groups ($n = 14$) were invited to respond to a survey questionnaire involving a set of 10 statements regarding their perceptions of the use of mobile TTS and ASR in this project. The statements asked participants if they believed that the use of TTS and ASR: (a) increased their motivation to learn about French pronunciation; (b) allowed them to become aware of some of their pronunciation problems; (c) allowed them to self-assess (decide whether their pronunciation was correct or incorrect); (d) made them feel more comfortable practicing pronunciation with the TTS/ASR apps than they would in front of other students; (e) made them feel more comfortable practicing pronunciation with the TTS/ASR apps than they would in front of the teacher; (f) contributed in their learning of French pronunciation; (g) helped

them to improve their French pronunciation; (h) should be used in other language courses to improve their pronunciation; (i) constituted great tools to learn pronunciation; and, finally, (j) was an enjoyable experience in their pronunciation class.

After the completion of the survey, the participants were scheduled for an oral interview with the researcher on a topic about which they had been informed previously: their experiences using TTS and ASR and their views on the strengths and weaknesses of using these technologies for learning purposes. The interviews, which lasted approximately 10 min each, were audio-recorded, transcribed, and later analyzed so that they could be categorized into the pertinent themes encompassed by the study.

The data from the survey questionnaire were analyzed using descriptive statistics, in which we established the mean values and associated standard deviations for each item under consideration. The interviews, on the other hand, were transcribed by two research assistants and the researchers and then analyzed according to the coding methods proposed by [44,45]. Based on these two authors' recommendation, we initially categorized the participants' responses based on the general questions that guided this research (i.e., learners' perception of TTS and ASR with respect to their strengths and weaknesses as pedagogical tools) and relevant subcomponents (e.g., the tools' ability to provide access to input and to promote output practices). In vivo coding was chosen as a first-cycle coding method in an attempt to more fully encapsulate participants' intended meanings. Codes were extracted verbatim from the transcripts and inserted into columns in a spreadsheet. Afterwards, pattern coding was used as a second-cycle coding method to group these verbatim codes by similarity to create themes, categories, and sub-categories, according to the main themes adopted in this study, namely, the participants' overall perceptions of L2 learning with TTS and ASR, and the pedagogical affordances that these technologies provided in terms of potential for input delivery, opportunities for output practice, and orthography-based predictions.

4. Results

4.1. Survey Analysis

The quantitative survey analysis aimed to answer the following question: What are learners' perceptions of the use of mobile TTS and ASR as pronunciation learning tools, based on a set of 10 pre-determined criteria? The data compiled via the survey questionnaire were analyzed via a simple mean calculation with associated standard deviation (descriptive statistics). Accordingly, to answer the first research question, means were used to measure the students' ratings of the statements adopted in the study. Table 1 illustrates the results for each of the 10 items selected. As established earlier, results closer to "5" correspond to "I strongly agree" with the statement presented. Given that both the TTS and ASR groups were presented with identical sets of survey questions and because of the significant overlap in the results (see Table 1), they are reported together to better highlight their similarities with regards to the 10 criteria adopted in the study.

The results in Table 1 show that the L2 French learners who participated in these two studies view the use of mobile TTS and ASR positively, with most selecting "agree" and "strongly agree" for all the items included in the survey. Consider for example, the items that closely relate to the goals of this research, particularly items #2, #6 and #7, which reinforce our hypotheses that the two technologies could be used for the development of phonological awareness (item #2), to learn about pronunciation features in general (item #6), and to help one's overall pronunciation skills (item #7). They also revealed that the two adopted technologies increased their motivation to learn about French pronunciation (item #1), and made them feel more comfortable practicing on their own than in front of other students and the teacher (items #4 and #5 respectively). Interestingly, most participants would like to extend TTS and ASR use to other language courses (item #8). It is interesting to observe that, even though TTS's ability to provide feedback is limited (i.e., in comparison with ASR, which spells out orthographically whatever is dictated), most participants rated this tool's ability to allow them to evaluate their own pronunciation very highly (item #3). Finally, most participants strongly agreed

that the two apps are great learning tools (item #9), and that their use contributed to their overall enjoyment in the process of learning the two target features of French pronunciation (item #10).

In sum, our analysis of the survey results shows that, overall, the participants evaluated positively their experience with these two mobile speech technologies and, more importantly, they find them useful, practical and helpful for their own learning.

Table 1. Survey results: Learners' perception of mobile TTS and ASR ([App] refers to the respective app used: TTS or ASR).

Questions	Text to Speech Synthesizer (TTS)		Automatic Speech Recognition (ASR)	
	Mean	SD	Mean	SD
1. [App] increased my motivation to learn about French pronunciation.	4.1	0.83	3.86	1.10
2. [App] allowed me to become aware of some of my pronunciation problems.	4.1	0.35	3.86	0.77
3. [App] allowed me to evaluate my own pronunciation (for example, to decide whether my pronunciation was correct or incorrect).	3.6	0.74	3.93	1.07
4. I felt more comfortable practicing pronunciation with [App] than I would in front of other students.	3.8	1.28	3.64	1.28
5. I felt more comfortable practicing pronunciation with [App] than I would in front of the teacher.	3.3	1.16	3.36	1.28
6. I learned more about French pronunciation using [App] than if I had not used it.	3.8	0.89	3.64	0.93
7. In general, [App] helped me improve my pronunciation.	3.6	0.74	4.14	0.77
8. I would like to use [App] in other language courses to improve my pronunciation.	4.5	0.53	4.21	0.80
9. [App] is a great tool to learn pronunciation.	4.3	0.71	3.86	0.95
10. I enjoyed using [App] in this study.	4.6	0.52	4.21	0.58

4.2. Interview Analysis

We now report the results of the qualitative components of the study, which asked the following question: what are the perceived strengths and weaknesses of using TTS/ASR for learning pronunciation in a second language environment, particularly with respects to the tool's ability to provide access to input, increase output opportunities, and develop learners' prediction skills? This question was addressed via a content analysis of open-ended oral interviews with the participants, using the protocols set forth by [44,45], as described earlier.

The answers to the set of open-ended questions (which started with: "According to your experience, what are the strengths and weaknesses of using TTS/ASR applications for learning/practicing pronunciation in an L2 French classroom?") were compiled into canonical quotes, each representing the general concepts conveyed in their responses. The participants' answers were initially compiled into two main categories reflecting their perceptions of the strengths (Table 2) and weaknesses (Table 3) of the TTS/ASR applications. We then divided the general topics into subcomponents that more directly reflected the goals of the study: whenever appropriate, answers were further categorized into one of the three themes included in the research questions: input (increased or enhanced exposure to French), output (increased oral practice of French), and prediction (learning about pronunciation via spelling)—see Table 4.

As was the case in the reporting of the survey results, we report our analysis of the participant interviews for both the TTS and ASR studies together because they utilized identical sets of interview questions. This will also allow us to highlight the similarities between the two studies, to point out the differences in perceptions related to each tool, and to illustrate their complementarity in terms of input, output, and prediction capabilities.

Table 2. Perceived strengths of TTS and ASR.

1. Increased Involvement Outside of Class (Learning and Practicing Pronunciation Anytime, Anywhere)	
TTS (*n* = 5)	ASR (*n* = 3)
a. " . . . as long as I had a Wi-Fi connection, I can just take my tablet and listen to it anywhere." b. "You can do it whenever you want, so . . . I used it when I was free." c. "Mostly at home because I only had time on the weekend." d. "It's on the phone so you can carry it around anywhere." e. "Instead of trying to track someone down to help you, uh, it could help a lot."	a. "It's good to have homework that you can take home and practice pronouncing on your own, instead of just in the lab. So, yeah. Especially when there's no one else, different exercises." b. "I don't have much time to practice or much people to practice with and I really liked having these weekly assignments and to be able to practice at home." c. "You can use it anywhere. Anytime like if you have a doubt you can always just take it out and start using it."
2. Usefulness for Any L2 Classroom	
TTS (*n* = 6)	ASR (*n* = 7)
a. "Yeah, I think it would be really useful. And if the sentences on the app match what we will learn in class, that would be so useful, especially for learning new vocabulary, because, it's much easier to learn vocabulary when you can hear how it sounds and you can make the connection that way." b. "Yes, I think it's a good tool . . . I think, yeah, it's actually better than the French lab, I think, probably."	a. "They should definitely implement that in the grammar classes because it's like you need to know how to pronounce things, like, it's critical." b. "As an additional tool, for sure! We don't have much time to speak with other people or even with the teacher because we have a lot of students in the class. It's difficult to practice the pronunciation. Actually, I think pronunciation is important. I don't know why. Because the emphasis . . . during like general French class, not only the pronunciation class."
3. Increased Confidence (to Speak French)	
TTS (*n* = 0)	ASR (*n* = 2)
N/A	a. "I get nervous when it's in person. So, it's definitely easier, and then I can get more comfortable, and I don't mess up so much, in person . . . you just get more confident." b. "It's good because . . . maybe if someone is shy to speak, or even with the teacher 'cause sometimes we feel a little bit down when trying to speak and to pronounce . . . and with this application we can kind of try to do any sound, open your mouth and exaggerate and everything. It's just you and the computer. You can do whatever you want. No one is going to see. No judgments."
4. Easy to Use/Usability and Portability	
TTS (*n* = 7)	ASR (*n* = 2)
a. "just how easy it is, you know, if you have a tablet, it's right there, a couple of clicks and you can start listening." b. " . . . the thing that I kind of liked about it . . . [is] that it's portable"	a. "It's really easy. You just have to press a button, and then it records. I liked it a lot, yeah." b. "I think it's very easy to use and it's very portable."

The initial topics analysis indicated that the majority of the participants (*n* = 16/18) found the use of mobile TTS and ASR for pronunciation learning useful and helpful (*n* = 8/8 for TTS and *n* = 8/10 for ASR). As illustrated in Table 2, some of the responses stressed the benefits of the pedagogical use of TTS and ASR, which were often perceived as tools that: (1) increase involvement outside of class by providing opportunities for learning and practicing pronunciation anywhere and whenever ("You can use it anywhere. Anytime like if you have a doubt you can always just take it out and start using it"); (2) are useful for any L2 classroom because they compensate for the lack of time to practice pronunciation skills in class ("As an additional tool, for sure! We don't have much time to speak with other people or even with the teacher because we have a lot of students during the class. It's difficult to practice the pronunciation"); (3) increase learners' confidence levels in speaking French, as one of the participants indicated: "I get nervous when it's in person. So, it's definitely easier, and then I can get more comfortable, and I don't mess up so much in person . . . you just get more confident"; and (4) are portable and easy to use ("It's really easy. You just have to press a button, and then it records.

I liked it a lot"). It is important to note that the overwhelming majority of the participants agreed that the apps would be a great supplemental tool for any L2 classes for practicing not only general oral skills but also learning new grammatical structures and vocabulary "because it's much easier to learn vocabulary when you can hear how it sounds and you can make the connection that way". Finally, participants felt that the use of mobile TTS and ASR contributed to building their autonomy as learners: "instead of trying to track someone down to help you, it could help a lot".

Table 3 illustrates some of the perceived weaknesses of the pedagogical use of mobile TTS and ASR, as utilized in this study. The participants reported a lack of accuracy for both TTS and ASR. For example, for ASR, one of participants stated a lack of accuracy in spelling out what he had just said ("I feel like it didn't pick up what I was saying exactly. Some bizarre words came up that were just completely unrelated to what I was trying to tell my phone"). For TTS, most participants noted the robotic and unnatural nature of the voices ("The only thing that was really a weakness for me, it's probably just the fact that it is an electronic voice. It's not, like, natural sound"). Another weakness reported was the two tools' inability to provide explicit corrective feedback, as the implicit feedback provide via sound (TTS) or orthography (ASR) was not always sufficient to inform the participants whether their pronunciation was correct, or to enable them to correct themselves: "Maybe just . . . even a link or something . . . even the IPA. . . . when I was getting to the thirteenth, the fourteenth [try] and I'm just like, 'I don't know how you want me to say it!'" Finally, a small number of students complained that the operation of the ASR program was time consuming ("You have to press it, and then record it, and sometimes it [takes a] very long time") and the TTS lacked interactivity ("for me the best part of learning French is communicating with people. [This app] is more listening than speaking").

Table 3. Perceived weaknesses of TTS and ASR.

1. Lack of Accuracy	
TTS (*n* = 5)	ASR (*n* = 6)
Pronunciation Accuracy (Different from Natural Voice):	Accuracy of Voice Recognition:
a. "The only thing is sometimes it makes weird sounds, like, automatic. You know, like, the computer voice thing? So, sometimes it's harder to understand . . . " b. " . . . but in some cases with pronunciation, for example, it sounded, like, really, it was coming from a computer . . . a few of the words were really just not pronounced in a way that I could even hear it." c. "Yes, I still feel like it's pronounced by a machine ... even in Chinese . . . a machine talking in Chinese . . . I feel like it's a bit weird . . . I'm not sure it's a "natural" reader." d. "It's like with the GPS . . . " e. "The only thing that was really a weakness for me, it's probably just, like, the fact that it is, like, an electronic voice. It's not, like, natural sound."	a. "Well, I don't know if it's my pronunciation that is bad, or whether it was the technology of the instrument, but I didn't think . . . some of the words that popped up to me weren't even in the ballpark of what I was saying. Some of them were similar; some of them were completely out." b. "I feel like it didn't pick up what I was saying exactly. Some bizarre words came up that were just completely unrelated to what I was trying to tell my phone." c. "Sentences. Maybe a negative point . . . I see when you try to speak with sentences they don't recognize it very well. One word it's okay but the same word in a sentence . . . I couldn't get it right." e. "it's good. But it was all frustrating 'cause sometimes I've been so off."
2. Lack of Explicit Corrective Feedback	
TTS (*n* = 3)	ASR (*n* = 5)
a. "Then again, watching TV and that in French will help too. It's on more or less the same level. It's, like, you learn how to pronounce it, but at the same time you don't get anyone correcting you." b. "I guess it would be difficult to do, but . . . having the person record and then balancing it between the two sound files and seeing, how the pronunciation is, . . . see how it compares." c. "Just thinking very hypothetically, maybe if there was a possibility to record your own pronunciation and then kind of hear how you sound after listening to your pronunciation."	a. "It would be nice if it gave us examples of the words. For example, if you say something wrong, it says 'Do you mean that?' 'This is what you pronounced . . . " b. "But sometimes I didn't know what to change, so I just said the same thing over and over." c. "I would have felt more motivated if I'd felt that I was learning something or improving . . . I didn't even know what I was doing wrong, because it didn't give any feedback." d. "Maybe just . . . even a link or something . . . even like the IPA. I don't know but when I was getting to the thirteenth, the fourteenth [try] and I'm just like 'I don't know how you want me to say it!'" e. "The one thing though would be to get the personal feedback, like after, and then like go through how to say each word. You say 'No...', what went wrong . . . "

Table 3. *Cont.*

3. Time Consuming	
TTS (*n* = 0)	ASR (*n* = 3)
N/A	a. "You have to press it, and then record it, and sometimes it [takes a] very long time." b. "It's very time-consuming to use. You say and then it pops up, and then you've got to delete, then get back. And it's a lot of manipulation of the device for very little reward." c. "What I'd like is not to have to keep pushing buttons all the time. Every time you pronounce something, you have to press back. It would be nice if it recorded things automatically and continuously."

4. Lack of Interactivity	
TTS (*n* = 2)	ASR (*n* = 0)
a. "But, uh, chatting with people probably helped a bit more." b. "Yeah, it's a good app to let me practice, for pronunciation . . . listening actually . . . but for me the best part of learning French is communicating with people. (The app) is more listening than speaking."	N/A

Finally, the excerpts in Table 4 illustrate the participants' perception of the pedagogical strengths and benefits of TTS/ASR for learning French pronunciation according to the three goals of pronunciation instruction (Dickerson's [16,17] trilogy of goals; the 3Ps): input (Perception), output (Production) and orthography (spelling-to-sound associations—Prediction).

In terms of input, the analysis of the interviews revealed that TTS is seen as a tool that helps to increase exposure to French. More specifically, it engages learners in extensive listening practice, helps to compensate for limited class time, allows them to practice oral comprehension, and promotes the learning of the pronunciation of unknown words. All the participants (*n* = 8/8) contributed with statements such as "I think it was really useful because some words that I don't know how to pronounce . . . I could read them and also hear how to pronounce them". In addition, some participants reported a perceived improvement of their pronunciation, which they explained by the fact that the app increased their exposure and access to the correct pronunciation model. Finally, two participants highlighted the benefits of the variability of the input provided by the TTS system, and noted that "the fact that you can pick the speed and the speaker is very helpful" and that they "have not just a female voice, but another voice [so] maybe some sounds would be easier to hear".

Regarding output, the majority of participants (*n* = 6/8) agreed that the use of ASR increased their practice of French pronunciation and contributed to learning. This can be exemplified with the following excerpts: "[App] gave me an idea of how my pronunciation improved since I started taking this class" and "My pronunciation is getting better now but before it wasn't very good. It's a way of practicing 'cause it's going to force you to practice".

In relation to prediction, both apps received positive comments from the participants. The TTS users highlighted the fact that the app allowed them to focus on the grapheme-to-phoneme associations by providing them with opportunities to "repeat [. . .] and [. . .] be able [. . .] to distinguish, whether I sound right or not", to "re-reading it and listening to it a few times, [. . .] it helps [with] catching the words and separating them", and to "be able to read and follow the words". The participants in the ASR study, on the other hand, appreciated the immediate automated feedback ("It gives you the answers, you can see") which allowed them to self-assess their pronunciation abilities ("[it] allowed me to become aware of some of my pronunciation problems" and "to evaluate my own pronunciation"). It is important to mention that the use of ASR increased participants' motivation to produce language, as illustrated by one of the participants' comments: "I like when the first time when I speak it gets it right. I'm like 'Yeah, I can do it'".

Table 4. Learners' perceptions of TTS and ASR according to the tools' ability to increase/enhance the input, and create opportunities for output practice and for the development of prediction skills (3Ps).

Input (TTS)	
Increased exposure to French (*n* = 7)	
a. "it's very useful for me particularly, because I find . . . oral comprehension to be the most difficult part of learning French . . . In a class situation we don't actually get a lot of listening practice . . . so just to hear sentences spoken really clearly, . . . and also using vocabulary that we're actually learning in class, you know, was very useful for me." b. "I think it was really useful because some words that I don't know how to pronounce, like, I could read them, and also hear how to pronounce them." c. "It was very helpful for me to hear. I found the pronunciation very good, because sometimes I would see a word and I wasn't sure, but hearing them over and over I found was very helpful, and I feel like over the five weeks my pronunciation improved a little bit." d. "If you have a text and you want to hear how someone else pronounces a word that you've never heard before . . . that's actually one thing that I liked about it because, like, there's sometimes words that I don't even know, like, know how to pronounce it . . . but if I hear at least someone else saying it, I can repeat it and I can understand, you know?"	
Input variability (*n* = 2)	
a. " . . . well, I think the fact that you can pick the speed and the speaker is very helpful. So if one speed is too much for you, you can bring it down." b. " . . . have not just a female voice, but another voice [so] that maybe some sounds would be easier to hear . . . 'cause [in] some pitches . . . [it] is harder to hear some sounds than others."	
Output (ASR)	
Increased practice of French, pronunciation improvement (*n* = 6)	
a. "My pronunciation is getting better now but before it wasn't very good. It's a way of practicing 'cause it's going to force you to practice." b. "I like when the first time when I speak it gets [it] right. I'm like 'Yeah, I can do it.'" c. "[app] helped me improve the pronunciation of the French sound in 'u'." d. "In general, [app] helped me improve my pronunciation." e. "I learned more about French pronunciation using [app] than if I had not used it." f. "[app] gave me an idea of how my pronunciation improved since I started taking this class."	
Prediction (TTS and ASR, via Orthography)	
Helped me learn about pronunciation via spelling	
TTS (*n* = 5)	ASR (*n* = 4)
a. " . . . you can follow along. Also, in the paper there, I can just see how my pronunciation is, how it is like . . . if it's strong enough towards, the actual application, or whatever . . . whatever they're saying, I can repeat it and I can distinguish whether I sound [right] or not." b. " . . . and then I realized, ok, I'm starting to get the gist of it and then, so, for the third, fourth, and fifth I started listening to it and then writing how it sounded after the first go, and then afterwards, like, re-reading it and listening to it a few times, so . . . , in that sense, yeah, it helps with catching the words and separating them." c. " . . . it will help with hearing how stuff is pronounced." d. "I can note . . . correct pronunciation . . . like if the teacher is around me. And . . . I can practice a little bit [before] a lesson." e. "Well, I find it, like, really, um, interesting that you can read and follow the words."	a. "[app] allowed me to become aware of some of my pronunciation problems." b. "[app] allowed me to evaluate my own pronunciation (correct/incorrect)." c. "It gives you the answers, you can see . . . Because sometimes when I said some words, one of them was right, and then one of them was wrong, and then when one was wrong I was pronouncing it wrong . . . " d. "I like when the first time when I speak it gets [it] right."

To summarize, the results of the interview analysis corroborate in most aspects what was observed in the quantitative analysis. Overall, participants have positive perceptions of TTS and ASR and they are willing to use them in their own learning. As anticipated, learners believe that while TTS has strong potential to increase and enhance the language input to which they are exposed, ASR offers endless opportunities for output by allowing students to practice the target language anytime, anywhere, and at their own pace. In addition to the benefits, participants reported weaknesses and limitations of the two speech technologies adopted, including their lack of accuracy and explicit feedback.

5. Mobilizing Instruction: Discussion

The purpose of this study was to examine students' perceptions of the use of two speech technologies, TTS and ASR, as pedagogical tools to enhance their learning of L2 French pronunciation

skills. Two research questions guided this investigation: by means of a survey questionnaire, the first involved the participants' ratings of a set of statements addressing key factors regarding their experience using TTS and ASR (e.g., motivation level, phonological awareness, ability for self-assessment, comfort to practice, pronunciation improvement, and overall enjoyment). The second question inquired about the participants' assessment of their experience using the two technologies, with a focus on their strengths and weaknesses and, more importantly, their ability to address the three competence elements that constitute what Dickerson [16,17] defines as the trilogy of goals (3Ps) for pronunciation instruction: perception (input), production (output) and prediction (spelling-to-sound associations).

With regards to the first question and the related 10 criteria to evaluate learners' perceptions of technology, the overall results suggest that the participants view the use of the two speech technologies positively for all items included in the survey. The participants' ratings of the statements indicate they believe that the mobile use of TTS and ASR: increased their motivation to learn French; allowed them to become aware of their pronunciation problems and to evaluate their performance; made them feel more comfortable to practice on their own, via their mobile devices, rather than in front of their peers and teachers; helped them improve their overall French pronunciation; and, also importantly, had a great level of enjoyment using the adopted mobile versions of TTS and ASR as learning tools.

Despite the lack of studies that examine learners' perceptions and the learning potential of TTS and ASR, these results are in line with previous research on the pedagogical use of technologies that foster learner-centred teaching [38,49,50] as a style of instruction that is "responsive, collaborative, problem-centred, and democratic in which students [. . .] decide how, what, and when learning occurs" [51] (p. 42). The learner-centered environment afforded by these two technologies, combined with other factors such as an overall belief that learning is indeed taking place (e.g., "My pronunciation is getting better now but before it wasn't very good. It's a way of practicing 'cause it's going to force you to practice"), may have contributed to an overall sense of responsibility for one's own learning and consequently accomplishment. This sense of responsibility combined with continuous hypothesis testing ([23] e.g., via regular practice with TTS and ASR) may explain why our participants had a high level of motivation to learn about pronunciation, to self-assess, and to feel more comfortable and satisfied with their overall learning experience.

The second research question addressed (1) learners' perception of the strengths and weaknesses of using the two speech technologies in their classes; and (2) their ability to provide access to input, to increase output opportunities, and to develop learners' prediction skills. Starting with the second part of the question, for ease of exposition, the results obtained corroborate our hypothesis that, from a learner's perspective, the combined use of TTS and ASR (particularly their reliance on orthography for input and output respectively) seem to address the three goals of pronunciation instruction proposed by Dickerson [16,17], namely the development of perception, production and prediction skills. While TTS provides learners with the necessary input for increased and enhanced exposure to target forms (*perception*; including noticing and focus-on-form activities, as recommended by [52]—see also [24]), ASR affords learners with multiple opportunities for output so that they can practice and test their hypotheses about what is being learned (*production*; see [53], who claims that when learners are pushed to produce language, they may notice a gap between what they intend to say and what their interlanguage allows them to say). The ability to *predict*, the third item in Dickerson's trilogy of goals, derives from the use of either or both tools, which help learners generalize pronunciation patterns based on the orthographic representation of what they hear (TTS) or say (ASR). As described in the preceding section, many of our respondents' statements reflect, indirectly, the sound-to-spelling and spelling-to-sound affordances of the two speech technologies which, in many cases, constituted instances of implicit feedback about the students pronunciation (e.g., "sometimes, when I said some words [. . .] and then when one was wrong, I was pronouncing it wrong", as ASR user concluded).

Let us now address the general question about learners' perception of the strengths and weaknesses of TTS and ASR as pedagogical tools to learn about French pronunciation. As was the case

in the quantitative survey, students also responded positively to the use of the mobile versions of TTS and ASR in most of the themes identified and selected for the analysis of the oral interviews. Focusing on three of these topics, they included the tools' pedagogical usefulness, their ability to increase involvement outside of class (e.g., learning and practicing pronunciation at their own convenience, anytime anywhere), and their ability to boost their confidence levels to practice and to speak. These patterns are usually found in the CALL literature, particularly when learners interact with a computer (i.e., without the presence of a human). These types of computer-mediated communication have shown to increase student motivation [54], which in turn decrease learner anxiety [55,56] and consequently increase learners' willingness to communicate in the target language [12]. In the context of our study, TTS and ASR act as veils that reduce the pressures that characterize face-to-face interactions [55,56].

The qualitative analysis also revealed some weaknesses of the pedagogical implementation of TTS and ASR for pronunciation instruction, as adopted in this study. While some of the critiques related to the intrinsic nature of each application or usability issues (e.g., the lack of interactivity in TTS use: "the best part of learning French is communicating with people"; the completion time of some ASR-based tasks "it's very time-consuming; you say and then [the spelling] pops up, and then you [have] got to delete, then get back"), others reflected current limitations of the apps. For example, a lack of accuracy in synthesizing text (TTS) or recognizing speech (ASR) was one of the most cited weaknesses of the two applications, particularly TTS's robotic speech and ASR's inability to always "pick up what [some of the participants were] saying exactly". Another commonly cited issue involved the two apps' ability to provide feedback. While participants recognize that some form of feedback is attainable ("sometimes I would see a word and I wasn't sure, but hearing them over and over [helped me] know how to pronounce it"; "sometimes, when I said some words [. . .] and then when one was wrong, I was pronouncing it wrong" for TTS and ASR respectively), the consensus was that these commercial applications were not designed with these goals in mind.

5.1. Limitations and Future Research

We must exercise caution in generalizing the results presented here, as the low number of participants, the context in which the data were collected, and the potential for a novelty effect undermine the generalizability of our findings. Let us discuss each one of these limitations, which we hope to address in further research.

One of the major limitations of our study is the modest participant sample size, which consisted of 23 participants who used one of the two target technologies, on separate occasions, and were then asked to engage in interviews to express their perceptions of and attitudes towards using TTS and ASR to learn features of French phonology. It would be interesting to investigate the pedagogical use of these technologies with a larger population of students using the two technologies at the same time. While TTS use would enhance the input in terms of quantity (increased exposure) and quality (e.g., via repetition, slow speech, and visual cues such as orthography), ASR would promote the practice of oral production skills, accompanied by immediate orthographic feedback (see forthcoming Section 5.2 for the pedagogical implementation of this idea).

Another less serious limitation, but worthy of consideration in a replication study, is the context in which the study was conducted, one in which French is used as a "second" language. Would similar results hold if the target were a "foreign" language, where exposure to and opportunities for interaction with the L2 are limited? According to [57] (p. 19), this constrained foreign language situation is characterized by a "pervasive L1 setting" that does not extend outside the classroom. In addition, it employs formal instruction that focuses on lexical and grammar instruction rather than on interaction. Finally, in these contexts, the source of L2 input in speech is impoverished and usually L1-accented, which can interfere with L2 learners' perception of phonetic and/or phonological details (e.g., the distinction between /u/ and /y/, which we targeted in this study). We hypothesize that, due to the restrictive nature of the foreign language context, the pedagogical and perceived benefits of TTS and ASR could be enhanced in this learning environment.

Finally, regarding learners' positive attitudes towards the pedagogical use of TTS and ASR, discussed in the previous section, we acknowledge the possibility that these perceptions are ephemeral, merely a reflection of the introduction of two novel technologies. This phenomenon has been observed in the CALL literature [58,59], potentially suggesting a novelty effect: "media do not directly influence learning," the instructional methods associated with the technology do [60] (p. 453). As an anonymous reviewer pointed out, once students are forced to use such technology, their motivation decreases. Only an extensive longitudinal study, conducted after the novelty factor has worn off, will be able to address the issue.

5.2. Pedagogical Implications

One way of circumventing some of the limitations uncovered by our research (e.g., the apps' ability to provide accurate forms consistently and the lack of options for explicit feedback) is by adopting mobile TTS and ASR applications as *complements* of the L2 classroom, in an environment that recognizes the importance of the ideal provider of "accurate" forms and "explicit feedback": the language teacher. In this scenario, we propose that the role of the teacher is to support the computer-assisted development of the 3Ps that encompass Dickerson's [16,17] trilogy of goals for pronunciation instruction, one in which the teacher has the role of a facilitator (along the lines of Brandl [61]), thus embracing many aspects of the learning experience (indicated by the non-linear arrows in Figure 3). Under this approach, Dickerson's trilogy of goals in a CALL/MALL context, illustrated earlier in Figure 1, now includes the language teacher with the important role of scaffolding and overseeing the language learning process, and ensuring that all five phases encompassed by Celce-Murcia et al.'s [1] framework for pronunciation instruction are obeyed. As shown in this study, while TTS and ASR may serve to assist in *perception* development (phases 1 and 2 in Celce-Murcia et al.'s framework, [1]), ASR has the potential to motivate learners to *produce* language in a controlled and/or guided manner (phases 3 and 4 in the same framework). Prediction, as previously described, results from the sound-to-spelling and spelling-to-sound capabilities afforded by the two technologies. The L2 teacher is then left with the role of assisting and supervising the successful completion of these stages by providing "accurate" forms and appropriate "feedback", and by encouraging students to practice the newly-acquired pronunciation features in a communicative way (phase five in Celce-Murcia et al.'s framework, [1]). A revised representation of our proposal is illustrated in Figure 3, now including the language teacher as the supervisor and overseer of the learning process.

The rationale behind the proposed teacher-supervised MALL-based approach to learning pronunciation is grounded on a body of research that highlights the importance of learning both in and outside the classroom [62,63]. One of the main benefits of this approach to learning is that it addresses the time constraint raised at the outset of this paper. By freeing up class time, teacher can then spend their limited resources on complementing their students' technology-enhanced, out-of-class learning experience (e.g., by providing explicit feedback and by engaging students in communicative practice). This idea was also articulated by many participants who recognized that classroom time is limited (e.g., "[ASR] as an additional tool, for sure! We don't have much time to speak with other people or even with the teacher because we have a lot of students in the class. It's difficult to practice the pronunciation"). Obviously, by engaging students in active and autonomous learning and extending the reach of the classroom, teachers might expect a relatively lower satisfaction by some students. As is generally the case in learner-centered education, the shift to a constructivist approach to learning is not always equated with full student satisfaction. For example, some students will have to work considerably more to improve their pronunciation abilities, while some will believe that, by doing so, they are "alleviating" the teacher's workload. We interpret these statements as epitomes of the pleasure-pain principle and, along the lines of Dupin-Bryant [51], as indications of a learning environment that is inherently learner-centered.
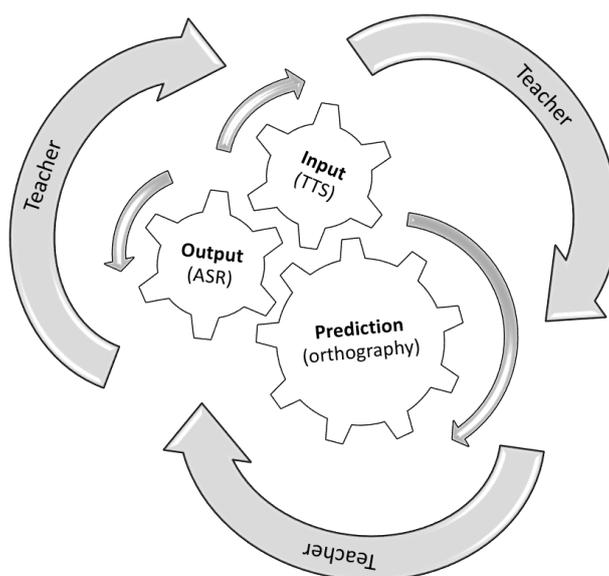
Figure 3. A proposal for the implementation of Dickerson's [16,17] trilogy of goals (3Ps) in a CALL/MALL context and the role of the instructor: to scaffold and oversee the process.

Author Contributions: Denis Liakin, Walcir Cardoso and Natallia Liakina contributed equally to the paper. Denis Liakin and Walcir Cardoso conceptualized the study, and Denis Liakin and Natallia Liakina collected the data. The paper was written by all three researchers.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Celce-Murcia, M.; Brinton, D.; Goodwin, J.; Griner, B. *Teaching Pronunciation: A Course Book and Reference Guide*; Cambridge University Press: Cambridge, UK, 2010; ISBN 0521729769.

2. Derwing, T.M.; Munro, M.J. Accent, intelligibility and comprehensibility: Evidence from four L1s. *Stud. Sec. Lang. Acquis.* **1997**, *19*, 1–16. [CrossRef]

3. Levis, J. Changing contexts and shifting paradigms in pronunciation teaching. *TESOL Quart.* **2005**, *39*, 369–378. [CrossRef]

4. Munro, M.J.; Derwing, T.M. Foreign accent, comprehensibility, and intelligibility in the speech of second language learners. *Lang. Learn.* **1999**, *49*, 285–310. [CrossRef]

5. Collins, L.; Muñoz, C. The Foreign Language Classroom: Current Perspectives and Future Considerations. *Mod. Lang. J.* **2016**, *100*, 133–147. [CrossRef]

6. Stockwell, G. Investigating learner preparedness for and usage patterns of mobile learning. *ReCALL* **2008**, *20*, 253–270. [CrossRef]

7. Pearson Student Mobile Device Survey. National Report: College Students. 2015. Available online: http://www.pearsoned.com/wp-content/uploads/2015-Pearson-Student-Mobile-Device-Survey-College.pdf (accessed on 18 January 2017).

8. UNESCO. Teaching and Learning: Achieving Quality for All. UNESCO, 2014. Available online: http://unesdoc.unesco.org/images/0022/002256/225660e.pdf (accessed on 12 January 2017).

9. Duman, G.; Orhon, G.; Gedik, N. Research trends in mobile assisted language learning from 2000 to 2012. *ReCALL* **2015**, *27*, 197–216. [CrossRef]

10. Liakin, D.; Cardoso, W.; Liakina, N. Learning L2 pronunciation with a mobile speech recognizer: French /y/. *CALICO J.* **2015**, *32*, 1–25.

11. Liakin, D.; Cardoso, W.; Liakina, N. The pedagogical use of mobile speech synthesis (TTS): Focus on French liaison. *Comput. Assist. Lang. Learn.* **2017**, *30*, 348–365. [CrossRef]

12. Reinders, H.; Pegrum, M. Supporting Language Learning on the Move. An evaluative framework for mobile language learning resources. In *Second Language Acquisition Research and Materials Development for Language*; Tomlinson, B., Ed.; Taylor & Francis: London, UK, 2016; pp. 116–141. ISBN 978-1441122933.

13. Stockwell, G. Commentary: Working with constraints in mobile learning: A response to Ballance. *Lang. Learn. Technol.* **2012**, *16*, 24–31.

14. Stockwell, G. Investigating an intelligent system for vocabulary learning through reading. *JALTCALL J.* **2013**, *9*, 259–274. Available online: http://journal.jaltcall.org/articles/9_3_Stockwell.pdf (accessed on 21 January 2017).

15. Van Praag, B.; Sanchez, H.S. Mobile technology in second language classrooms: Insights into its uses, pedagogical implications, and teacher beliefs. *ReCALL* **2015**, *27*, 288–303. [CrossRef]

16. Dickerson, W. Prediction in teaching pronunciation. In *The Encyclopedia of Applied Linguistics*; Chapelle, C., Ed.; Blackwell: Hoboken, NJ, USA, 2013; ISBN 978-1405194730.

17. Dickerson, W. Using orthography to teach pronunciation. In *The Handbook of English Pronunciation*; Reed, M., Levis, J., Eds.; Wiley Blackwell: Chichester, UK, 2015; pp. 488–503. ISBN 978-1-118-31447-0.

18. Ellis, N.; Collins, L. Input and second language acquisition: The roles of frequency, form, and function introduction to the special issue. *Mod. Lang. J.* **2009**, *93*, 329–335. [CrossRef]

19. Færch, C.; Kasper, G. The role of comprehension in second-language learning. *Appl. Linguist.* **1986**, *7*, 257–274. [CrossRef]

20. Chomsky, N.; Lasnik, H. The theory of principles and parameters. In *Syntax: An International Handbook of Contemporary Research*; von Stechow, J., Jacobs, A., Sternefeld, W., Vennemann, T., Eds.; de Gruyter: Berlin, Germany, 1993; ISBN 978-3110142631.

21. Flege, J.E. Second language speech learning: Theory, findings, and problems. In *Speech Perception and Linguistic Experience: Theoretical and Methodological Issues in Cross-Language Speech Research*; Strange, W., Ed.; York Press: Timonium, MD, USA, 1995; pp. 233–277. ISBN 978-0912752365.

22. Yanguas, I. Task-based oral computer-mediated communication and L2 vocabulary acquisition. *CALICO J.* **2012**, *29*, 507–531. [CrossRef]

23. Swain, M. The output hypothesis and beyond: Mediating acquisition through collaborative dialogue. In *Sociocultural Theory and Second Language Learning*; Lantof, J., Ed.; Oxford University Press: Oxford, UK, 2000; pp. 97–114. ISBN 978-0194421607.

24. Nation, P.; Newton, J. *Teaching ESL/EFL Listening and Speaking*; Routledge: New York, NY, USA, 2009; ISBN 978-0415989701.

25. Swain, M. Focus on form through conscious reflection. In *Focus on Form in Classroom Second Language Acquisition*; Doughty, C., Williams, J., Eds.; Cambridge University Press: Cambridge, UK, 1998; ISBN 978-0521625517.

26. Follick, M. *The Case for Spelling Reform: With a Foreword by Sir William Mansfield Cooper*; Manchester University Press: Manchester, UK, 1965; ISBN 978-0273419792.

27. Tranel, B. *The Sounds of French: An Introduction*; Cambridge University Press: Cambridge, UK, 1987; ISBN 978-0521304436.

28. Cardoso, W.; Collins, L.; White, J. Phonological Input Enhancement via Text-to-Speech Synthesizers: The L2 Acquisition of English Simple Past Allomorphy. In Proceedings of the American Association of Applied Linguistics Conference, Boston, MA, USA, 24–27 March 2012.

29. Cardoso, W.; Smith, G.; Garcia Fuentes, C. Evaluating text-to-speech synthesizers. In *Critical CALL—Proceedings of the 2015 EUROCALL Conference, Padova, Italy*; Helm, F., Bradley, L., Guarda, M., Thouësny, S., Eds.; Research-publishing.net: Dublin, Ireland, 2015; pp. 108–113.

30. Soler-Urzua, F. The Acquisition of English /I/ by Spanish Speakers via Text-to-Speech Synthesizers: A Quasi-Experimental Study. Master's Thesis, Concordia University, Montreal, QC, Canada, 2011.

31. Kirstein, M. Universalizing Universal Design: Applying Text-to-Speech Technology to English Language Learners' Process Writing. Ph.D. Dissertation, University of Massachusetts, Boston, MA, USA, 2006.

32. Guclu, B.; Yigit, S. 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. Human.* **2015**, *4*, 31–33.

33. Proctor, C.P.; Dalton, B.; Grisham, D.L. Scaffolding English language learners and struggling readers in a universal literacy environment with embedded strategy instruction and vocabulary support. *J. Lit. Res.* **2007**, *39*, 71–93. [CrossRef]

34. Couper, M.; Berglund, P.; Kirgis, N.; Buageila, S. Using Text-to-speech (TTS) for Audio Computer-assisted Self-interviewing (ACASI). *Field Methods* **2016**, *28*, 28–95. [CrossRef]

35. Cucchiarini, C.; Neri, A.; Strik, H. Oral Proficiency Training in Dutch L2: The Contribution of ASR-based Corrective Feedback. *Speech Commun.* **2009**, *51*, 853–863. [CrossRef]

36. Hardison, D. Generalization of computer-assisted prosody training: Quantitative and qualitative findings. *Lang. Learn. Technol.* **2004**, *8*, 34–52.

37. Hirata, Y. Computer assisted pronunciation training for native English speakers learning Japanese pitch and durational contrasts. *Comput. Assist. Lang. Learn.* **2004**, *17*, 357–376. [CrossRef]

38. McCrocklin, S. Pronunciation Learner Autonomy: The Potential of Automatic Speech Recognition. *System* **2016**, *57*, 25–42. [CrossRef]

39. Neri, A.; Mich, O.; Gerosa, M.; Giuliani, D. The effectiveness of computer assisted pronunciation training for foreign language learning by children. *Comput. Assist. Lang. Learn.* **2008**, *21*, 393–408. [CrossRef]

40. Strik, H.; Truong, K.; Wet, F.; Cucchiarini, C. Comparing different approaches for automatic pronunciation error detection. *Speech Commun.* **2009**, *51*, 845–852. [CrossRef]

41. Strik, H.; Colpaert, J.; van Doremalen, J.; Cucchiarini, C. The DISCO ASR-based CALL system: Practicing L2 oral skills and beyond. In *Proceedings of the Conference on International Language Resources and Evaluation (LREC 2012)*; European Language Resources Association (ELRA): Istanbul, Turkey, 2012; pp. 2702–2707. Available online: http://www.lrec-conf.org/proceedings/lrec2012/pdf/787_Paper.pdf (accessed on 29 January 2017).

42. Stockwell, G. Using mobile phones for vocabulary activities: Examining the effect of the platform. *Lang. Learn. Technol.* **2010**, *14*, 95–110.

43. Joseph, S.; Uther, M. Mobile devices for language learning: Multimedia approaches. *Res. Pract. Technol. Enhanc. Learn.* **2009**, *4*, 7–32. [CrossRef]

44. Saldaña, J. *The Coding Manual for Qualitative Researchers*; Sage: Los Angeles, CA, USA, 2009; ISBN 978-1446247372.

45. L'Écuyer, R. L'analyse de contenu: Notion et étapes. In *Les Méthodes de la Recherche Qualitative*; Deslauriers, J.-P., Ed.; Presses de l'Université du Québec: Sainte-Foy, QC, Canada, 1990; pp. 54–63. ISBN 978-2760504295.

46. Chapelle, C. *Computer Applications in Second Language Acquisition: Foundations for Teaching, Testing, and Research*; Cambridge University Press: Cambridge, UK, 2001; ISBN 9780521626460.

47. Chapelle, C. Using Mixed-Methods Research in Technology-Based Innovation for Language Learning. Proceedings of Innovative Practices in Computer Assisted Language Learning Conference, University of Ottawa, Ottawa, ON, Canada, 26–27 April 2012.

48. Cardoso, W. Learning a foreign language with a learner response system: The students' perspective. *Comput. Assist. Lang. Learn.* **2011**, *24*, 393–417. [CrossRef]

49. Pillay, H. Understanding learner-centredness: Does it consider the diverse needs of individuals? *Stud. Contin. Educ.* **2002**, *24*, 93–102. [CrossRef]

50. Weimer, M. *Learner-Centered Teaching*, 2nd ed.; Jossey Bass: San Francisco, CA, USA, 2013; ISBN 978-1118119280.

51. Dupin-Bryant, P.A. Teaching Styles of Interactive Television Instructors: A Descriptive Study. *Am. J. Distance Educ.* **2004**, *18*, 39–50. [CrossRef]

52. Doughty, C.; Long, M. Optimal psycholinguistic environments for distance foreign language learning. *Forum Int. Dev. Stud.* **2003**, *23*, 35–75.

53. Swain, M. The output hypothesis: Theory and research. In *Handbook of Research in Second Language Teaching and Learning*; Hinkel, E., Ed.; Lawrence Erlbaum: Mahwa, NJ, USA, 2005; pp. 471–483. ISBN 978-0805841817.

54. Ducate, L.; Lomicka, L. Podcasting: An Effective Tool for Honing Language Students' Pronunciation? *Lang. Learn. Technol.* **2009**, *13*, 66–86.

55. Arnold, N. Reducing foreign language communication apprehension with computer-mediated communication: A preliminary study. *System* **2007**, *35*, 469–486. [CrossRef]

56. Baralt, M.; Gurzynski-Weiss, L. Comparing learners' state anxiety during task-based interaction in computer-mediated and face-to-face communication. *Lang. Teach. Res.* **2011**, *15*, 201–229. [CrossRef]

57. Best, C.T.; Tyler, M. Nonnative and second-language speech perception: Commonalities and complementarities. In *Second-Language Speech Learning: The Role of Language Experience in Speech Perception and Production*; Bohn, O.S., Munro, M., Eds.; John Benjamins: Amsterdam, The Netherlands, 2007; pp. 13–34. ISBN 978-9027219732.

58. Nikolova, O. Effects of Students' Participation in Authoring of Multimedia Materials on Student Acquisition of Vocabulary. *Lang. Learn. Technol.* **2002**, *6*, 100–122.

59. Warschauer, M. Motivational aspects of using computers for writing and communication. In *Telecollaboration in Foreign Language Learning: Proceedings of the Hawai'i Symposium*; Warschauer, M., Ed.; University of Hawai'i, Second Language Teaching and Curriculum Center: Honolulu, HI, USA, 1996; ISBN 978-0824818678.

60. Clark, R.E. Reconsidering research on learning from media. *Rev. Educ. Res.* **1983**, *53*, 445–459. [CrossRef]

61. Brandl, K. Integrating internet-based reading materials into the foreign language curriculum: From teacher- to student-centered approaches. *Lang. Learn. Technol.* **2002**, *6*, 87–107.

62. Engin, M. Extending the flipped classroom model: Developing second language writing skills through student-created digital videos. *J. Scholarsh. Teach. Learn.* **2014**, *14*, 12–26. [CrossRef]

63. McCombs, B.L. Assessing the role of educational technology in the teaching and learning process: A learner-centered perspective. In *Proceedings of the Secretary's Conference on Educational Technology: Measuring the Impacts and Shaping the Future*; (ERIC Document Reproduction Service No. ED452830); Department of Education: Washington, DC, USA, 2000. Available online: http://files.eric.ed.gov/fulltext/ED452830.pdf (accessed on 10 January 2017).