

The effects of imitation and synchronization on the pronunciation of selected phonemes in L2 English and German: a pilot study

Wenxun Fu, Barbara Kühnert, Claire Pillot-Loiseau, Simone Falk

Laboratory of Phonetics and Phonology, UMR 7018, CNRS - Paris 3 Sorbonne Nouvelle University, France
(wenxun.fu, barbara.kuhnert, claire.pillot, simone.falk)@sorbonne-nouvelle.fr

ABSTRACT

This study focuses on investigating possible effects of synchronous speech, an experimental version of joint speech [3], on L2 pronunciation at the segmental level. While repetition and imitation are traditionally used in pronunciation teaching and learning of L2 phonetic and phonological acquisition, synchronous speech has been seldom studied in an L2 learning environment. What are the L2 linguistic aspects that synchronous speech would influence? Is there any effect of L2 phonetic convergence found when learners are speaking and listening at the same time? We studied the effects of synchronization and imitation on the acquisition of eight phonemes in L2 English and German that are known to be problematic for French learners. A series of acoustic analysis revealed that while some acoustic parameters such as formant frequencies and vocalic duration improved in both speech practices, changes brought on by synchronous speech of consonantal acoustic parameters were more subtle to determine.

Keywords: joint speech, synchronous speech, L2 learning, imitation, phonetic convergence.

1. INTRODUCTION

Often observed among speakers in collective activities, joint speech is found principally in such occasions as prayer meetings and civil manifestations. Joint speech has received increased attention recently for its contribution to the remediation of fluent speech in patients suffering from non-fluent aphasia [13] and stuttering [9]. However, despite its clinical popularity, joint speech has been rarely considered as a method of pronunciation training in foreign language (L2) learning. Meanwhile, quite a number of L2 pronunciation training approaches include imitating native-speaker pronunciation models, a traditional method whose efficacy has been approved both theoretically and practically. Imitation of speech sounds has been shown to involve all levels of spoken language, and adaptation to the interlocutor on the acoustic-phonetic level is defined as phonetic convergence, which is considered as a driving

mechanism in the acquisition of the phonology and phonetics of a L2 [10].

However, it should be noted that phonetic convergence in natural conversational settings is typically subtle. Unintentional imitative changes have been shown to be weaker than voluntary imitative ones [12]. Therefore, when we want to make use of phonetic convergence to benefit L2 speech learning, particularly in training focusing on the authenticity of pronunciation, a task of repetition with clear instruction of imitation could be more efficient. Independent of willingness, the degree of imitation also appears to be strongly correlated to the closeness of time between the input and the production of speech. For example, increased phonetic convergence effects have been shown in close shadowing when compared with delayed shadowing tasks [7]. Since joint speech, and its experimental pendant synchronous speech [4], can be considered as an extreme case in which input and production occur (almost) simultaneously, we were interested in the question of whether phonetic convergence effects would be maximized in synchronous speech, or on the contrary, be suppressed due to the lack of time for perceptive learning. The goal of this study was to investigate potential changes in the relevant acoustic features of selected acoustic vowel and consonant targets during synchronous speech and voluntary imitation in an L2 learning environment. Furthermore, we also wanted to evaluate whether the effects of the two different production tasks on the learners' performance differ according to whether the target language is familiar or unfamiliar.

2. METHODS

2.1. Participants

A group of 13 female bachelor students with an age range of 18-22 years, (mean age 20, SD 1.3), all native French speakers, participated in both the imitation and synchronous speech experiment. Except for three participants who indicated that their

level corresponded to A1 of the CEFR¹, all others had never learned German. All 13 participants had learned English at school, but considered their level to be of different proficiency (ranging from A1 to C2 of the CEFR) according to the analysis of their self-assessment questionnaires. Regarding foreign accents ratings, 4 out of 13 believed they had a relatively strong accent while others judged their accent as moderate.

2.2. Stimuli

For a pre- and post-test in English, the following words were chosen: *part*, *peace*, *sad*, *pear*, *boat*, *lit*, *feat*, *but*, and *met*. For the imitation and synchronous speech experiment, eight phonemes, /ɪ/, /i:/, /s/ and /θ/ in English, and /ɪ/, /i:/, /b/ and /p/ in German, were selected as target phonemes. As French lacks the distinction between /ɪ/ and /i:/, speakers have a tendency to produce both sounds similar to the French vowel /i/.

The two vowels are also known to be often distinguished only by duration rather than by both duration and spectral values, a pronunciation mistake found in learners of English of different language backgrounds [5], [6]. /θ/ is frequently substituted by /s/ among native French speakers [11]. Unlike in English and German, /b/ in word-initial position is voiced in French, while /p/ is voiceless in all three languages but only aspirated in English and German.

For each pair of target phonemes, seven minimal pairs were selected (see Table 1) and included in a carrier sentence (English: “Did Toby say “*target word*”?”, German: “Sagte Tina “*target word*”?”). A female native English and native German speaker recorded 28 sentences with the target minimal pairs, plus 14 further sentences serving as either before-test training or as distracters. For both conditions, the 28 test and 10 distracter sentences were randomized. For the imitation experiment, three beeps were inserted with an interval of 750ms each between sentences. Intervals between the end of a sentence and the first of three beeps were also 750ms. The intervals between the end of the third beep and the start of the next sentence were 5s. For the synchronous speech experiment, the 38 sentences were first reproduced resulting in 76 sentences. Again, the two repetitions of each sentence were presented following and preceding three beeps with 750ms intervals. The first repetition served as an example sentence, the second repetition as the synchronization sentence. The deliberately repetitive rhythm of the experiment was created to help participants to anticipate the beginning

of synchronous speech and thus to better synchronise their speech with the pronunciation model.

Table 1: Examples of target minimal pairs in English and German.

English	/ɪ/ <i>bit</i>	/i:/ <i>beat</i>	/s/ <i>sing</i>	/θ/ <i>thing</i>
German	/ɪ/ <i>Zinn</i>	/i:/ <i>ziehn</i>	/b/ <i>Bier</i>	/p/ <i>Pier</i>

2.3. Procedure and Data Analysis

Before and after the experiment in English, the 13 subjects were asked to read the 10 words as a pre and post-test. Each subject participated in four main tests: imitation in English, imitation in German, synchronous speech in English and synchronous speech in German. Each test lasted around 8 minutes. The experiments in English and German took place on different days in order to minimize potential cross-linguistic influences.

The recordings were segmented and phone-labelled manually in Praat [1]. To analyse and compare the pronunciation quality of the learners’ productions, average values of different acoustic parameters were extracted. The formant values F1 and F2, and duration were analysed for /ɪ/ and /i:/ in both English and German. For /s/ and /θ/ in English, the spectral centre of gravity (COG) and duration were extracted. For /b/ and /p/ in German, VOT was measured. All duration measurements were normalized relative to word duration. For each acoustic parameter, a two-way ANOVA was calculated in R Studio². The two independent factors were *Condition* (imitation and synchronous speech) and *Phoneme* (phoneme1 and phoneme2). The level of significance was set at 0.05.

3. RESULTS

3.1. Vowels /ɪ/ and /i:/

The analysis confirmed immediate positive effects of imitation and synchronous speech on the articulation of the vowels /ɪ/ and /i:/ in both languages. As shown in Figure 1, the native French subjects did not produce English /ɪ/ and /i:/ distinctively in the baseline test (F1: $F_{(1,24)}=0.016$, $p=0.899$; F2: $F_{(1,24)}=0.002$, $p=0.996$). The participants, however, showed a prominent distinction between /ɪ/ and /i:/ in both English and German in both experimental conditions (Figures 2, 3). The temporal and spectral contrasts were nevertheless less prominent than those produced by the native speakers. The vowel productions in German, the language unknown to the subjects, showed a higher variability, when compared

¹ Common European Framework of Reference for Languages: Learning, teaching, assessment

²Rstudio Team (2015). Integrated Development for R. Rstudio, Inc., Boston, MA URL :<http://www.rstudio.com/>.

to their English counterparts. In the experiments of both languages, there is no significant difference between the productions of /ɪ/ and /i:/ in synchronous speech and those in imitation.

Figure 1: Plot F1/F2 of /ɪ/ and /i:/ in the English baseline test. Red: productions of the native speaker. Green and purple: productions of the participants.

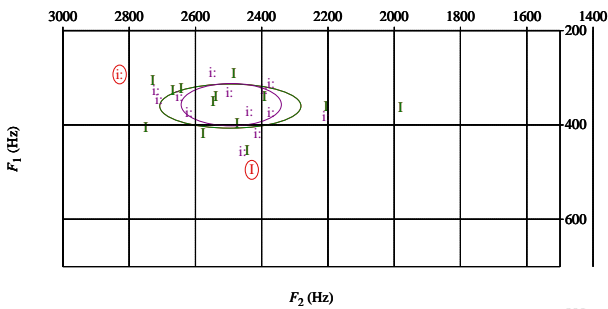


Figure 2: Plot F1/F2 of /ɪ/ and /i:/ in English. Red: productions of the native speaker. Green and blue: productions of the participants in imitation. Purple and black: productions of the participants in synchronous speech.

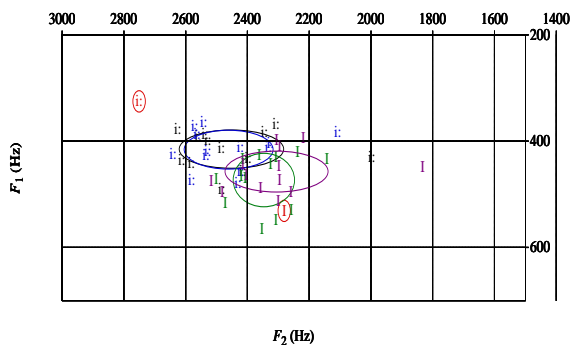
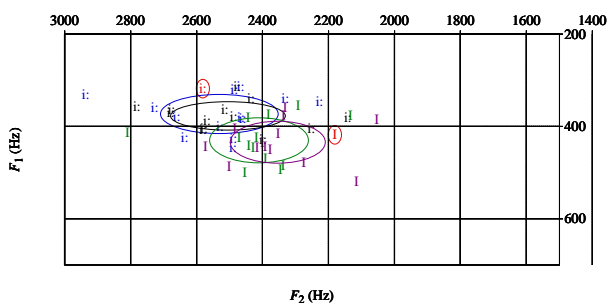


Figure 3: Plot F1/F2 of /ɪ/ and /i:/ in German. Red: productions of the native speaker. Green and blue: productions of the participants in imitation. Purple and black: productions of the participants in synchronous speech.



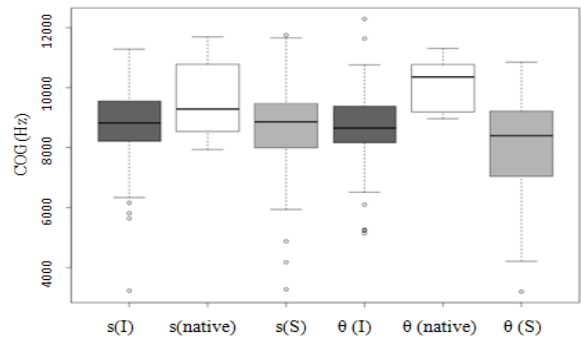
3.2. Consonants

3.2.1. /s/ and /θ/ in English

Analysis of both duration and spectral COG suggest that the French participants generally have difficulty in producing /θ/. The average COGs of /s/ and /θ/

produced by the English native speaker were 9674Hz and 10076Hz, respectively, and the duration of /θ/ was marginally shorter than the one of /s/, with relative duration being 38.3% and 44.9% respectively. As can be seen in Figure 4, the French participants in both conditions produced /θ/ with a lower COG than the native speaker. The two-way ANOVA did not reveal significant differences between the productions of /s/ and /θ/, neither with respect to the spectral energy ($F_{(1,50)}=2.327, p=0.134$), nor with respect to duration ($F_{(1,50)}=0.36, p=0.85$). Although a larger variability of COG of both consonants in joint speech seemed to be evident in Figure 4, once again, there was no statistical difference between the two conditions ($F_{(1,50)}=2.528, p=0.118$).

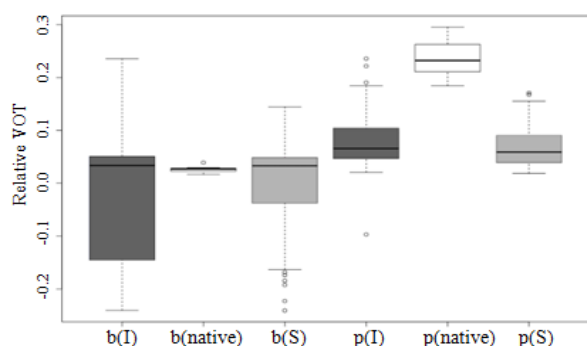
Figure 4: Median and variability of COG of /s/ and /θ/ produced by the English native speaker (native), subjects in imitation (I) and synchronous speech (S).



3.2.2. /b/ and /p/ in German

The statistical analysis showed a significant distinction between the two plosives ($F_{(1,50)}=43.683, p=2.92e-08$), but the difference between the two conditions ($F_{(1,50)}=0.103, p=0.749$) was insignificant; and there was no significant interaction ($F_{(1,50)}=0.417, p=0.240$). Four subjects produced /b/ in the German test sentences still with a voicing lead in all words in both experimental conditions, while several other speakers showed instances of pre-voicing in individual items (six speakers voiced /b/ in synchronous speech and seven in imitation). The duration of aspiration of all learners was considerably shorter than that of the native speaker. The results are summarized in Figure 5.

Figure 5: Median and variability of relative VOT of /b/ and /p/ produced by the German native speaker (native), subjects in imitation (I) and synchronous speech (S).



4. DISCUSSION

Generally speaking, the recordings of native speakers have facilitated the speech production of our subjects. All learners distinguished /l/ from /i:/, and /b/ from /p/, but not /s/ from /θ/. The effects of imitation and synchronous speech on the quality of the pronunciation of the eight target phonemes were, however, not significantly different. Even though several subjects have remarked that they felt more at ease and better completed the tasks in German in synchronous speech than in imitation, their productions in the synchronous conditions did not reflect the same positive effects, at least not at the segmental level.

The absence of the superiority of synchronous speech in the improvement of segmental pronunciations could be explained by two main reasons. Above all, a minimum amount of time would need to pass to allow external inputs to bring any effect on speech production at the segmental level. Phonemic hypothesis associated with articulatory objects should necessarily generate before the motor command is activated, and therefore the condition in which speakers synchronize their speech with others cannot exert any immediate effect on the adaptation of pronunciation at the segmental level.

Second, semantically, the instruction of synchronising one's own speech with the recording would have been translated by the subjects as "to start and finish the sentence at the same time as the recording". This, however, would have made them concentrate more on suprasegmental details than on segmental ones. The instruction of imitation, on the contrary, would have been understood as "to try to be similar" to the recorded sounds, and therefore forced the subjects to pay more attention to the phonetic characteristics of the sounds in order to meet the articulatory objectives.

Meanwhile, the lack of familiarity of the carrier sentences, specifically in German, would have prevented the French participants from synchronizing their speech with the model sounds at exactly the same time. Joint speech and synchronized speech as its experimental form [4], have always been conducted in conditions where all the speakers know rather well what they would say together. The speech in which a speaker constantly plays the role of « follower », even if it is subtle, would be more similar to *shadowing* or immediate repetition. Since our study did not measure the delay between the onset of the recording and the onset of the production of the participants, we could not determine whether the participants produced strict synchronous speech or whether they rather unconsciously closely followed the recording. Future studies could improve the experimental set-up and familiarize the subjects beforehand with the speech to synchronize.

5. CONCLUSION

Imitation, repetition and synchronous speech are all similar and different forms of speech in terms of the sensory-motor mechanisms they involve. In the case of phonetic convergence, sensory-motor interactions should permit an adaptation of articulatory and acoustic procedural knowledge to phonetic characteristics of the output. It seems from our study that in synchronous speech, effects of phonetic convergence and perceptive learning are difficult to be justified. This may explain why joint speech and synchronous speech have seldom been used as methods of improving pronunciation at the segmental level in L2 research. Our results are also coherent with other experimental and neuronal evidence that synchronous speech could be particularly facilitated by such suprasegmental cues as fundamental frequency, envelop amplitude [2] and rhythm [8].

Concerning methods of improving L2 pronunciation, our study confirmed positive effects of volunteer imitation on the accuracy of segmental pronunciation. However, to correctly produce foreign phonemes that are potentially problematic, training session of volunteer imitation should last longer than those provided in our experiment. Even though our study was based on a tightly controlled laboratory experiment, and hence a much more artificial environment than an actual language teaching class, we think that the implications of synchronous speech for pedagogical purposes cannot yet be excluded. Synchronous speech could still serve in L2 learning, but it would be important to distinguish its effects from that of imitation.

6. REFERENCES

- [1] Bersma, P., Weenink, D. 2001. Praat speech processing software. Institute of Phonetics Sciences of the University of Amsterdam. <http://www.praat.org>.
- [2] Cummins, F. 2009. Rhythm as entrainment: the case of synchronous speech. *Journal of Phonetics* 37, 16 – 28.
- [3] Cummins, F. 2013. Joint speech: the missing link between speech and music? *Percepta: J. Musical Cogn.* 1, 17-32.
- [4] Cummins, F. 2018. Comparing joint speech and synchronous speech: what happens when we added more speakers? *Proceedings of the 26th AIAI Irish Conference on Artificial Intelligence and Cognitive Science*, Trinity College, Dublin, Ireland, December 6-7th, 2018, 14-20.
- [5] Flege, J. E., Bohn, O. S. 1989. The perception of English vowels by native speakers of Spanish. *The Journal of the Acoustical Society of America* 85(S1), S85-S85.
- [6] Flege, J. E., Munro, M. J., MacKay, I. R. 1995. Effects of age of second-language learning on the production of English consonants. *Speech Communication* 16(1), 1-26.
- [7] Goldinger, S. D. 1998. Echoes of echoes? An episodic theory of lexical access. *Psychological Review* 105(2), 251-279.
- [8] Jasmin, K.M., McGettigan, C., Agnew, Z.K., Lavan, N., Josephs, O., Cummins, F. Scott, S. K. 2016. Cohesion and Joint Speech: Right Hemisphere Contributions to Synchronized Vocal Production. *Journal of Neuroscience* 6, 4669–4680.
- [9] Kalinowski, J., Saltuklaroglu T. 2003. Choral speech: the amelioration of stuttering via imitation and the mirror neuronal system. *Neurosci Biobehav Rev* 27, 339–47.
- [10] Nguyen, N., Delvaux, V. 2015. Role of imitation in the emergence of phonological systems. *Journal of Phonetics* 53, 46-54.
- [11] Picard, M. 2002. The differential substitution of English/θ ð/in French: The case against under specification in L2 phonology. *Linguisticae Investigationes* 25(1), 87-96.
- [12] Sato, M., Grabski, K., Garnier, M., Granjon, L., Schwartz, J. L., Nguyen, N. 2013. Converging toward a common speech code: imitative and perceptuo-motor recalibration processes in speech production. *Frontiers in Psychology* 4, 422.
- [13] Zumbansen, A., Peretz, I., Hébert, S. 2014. Melodic intonation therapy: back to basics for future research. *Frontiers in Neurology*, 5, 7.

7. ACKNOWLEDGEMENTS

This work is part of the program “Investissements d’Avenir” overseen by the French National Research Agency, ANR-10-LABX-0083 (Labex EFL).