Sources of variation of articulation rate in native and non-native speech: comparisons of French and German

Jürgen Trouvain, Bernd Möbius

Department of Computational Linguistics and Phonetics, Saarland University, Germany

`trouvain|moebius@coli.uni-saarland.de`

Abstract

Speech tempo including articulation rate is often considered as a good predictor in the diagnosis of foreign language proficiency and its comprehension. In this study we investigate various sources of variation of articulation rate such as the L2 proficiency level, individual tempo habits in L1 and L2, and more extensive exposure to native speech. In addition, we also discuss the difficulty of defining the most informative unit for rate metrics which allows comparisons between French and German. The materials used are French and German read sentences, produced as L1 and L2 speech. In contrast to other studies individual habits of articulation rate in the L1 was only partially observed in the corresponding L2 data (a slow L1 speaker does not necessarily articulate slowly in the L2). The convergence of most French learners to the German model speakers shows the advantage of having additional input for phonetic exercises. The fastest German learners also converge to the rather slow French model speaker.

Index Terms: articulation rate, L1, L2, convergence, individuality, French, German

1. Introduction

Speech tempo is often considered as a good predictor for various important concepts in the diagnosis of second and foreign language (henceforth L2) including the level of

- L2 proficiency [1],
- intelligibility and comprehension [2], and
- perceived foreign accent [2].

Articulation rate as the key component of speech tempo would thus be an optimal and easy-to-handle indicator of the level of the *spoken* language learning process.

It can be seen as established that *on average* speech in the first language (L1) is articulated faster than L2 speech, be it that the same speakers are faster in their L1 than in their L2 [3, 4, 5] or that L1 speakers are faster than L2 speakers in a given language [4, 6, 5]. Therefore the quantified tempo, whatever be the metrics, should reflect this difference between L1 and L2 speech. In addition, L2 speakers with a lower level of L2 proficiency will in general be slower than L2 speakers with a higher level of L2 proficiency [7].

Another factor of variability is the fact that languages show different measured rates. When we compare German with French, French usually shows 'faster' syllabic rates than German. One could argue that articulation in French is actually *produced* faster because it is *perceived* as faster than German – by German listeners. However, this argument can be applied in the other direction as well: articulation of German is also perceived as faster than 'usual' – by French listeners.

Cross-language studies on speech tempo either compare L1 speakers of one language with L1 speakers of another language (ideally with comparable text material), or L1 speakers of a given language with L2 speakers of the same language (ideally with identical text material).

A further source of variation lies in the fact that individual speakers differ in their tempo [10]. Some speakers articulate faster than others, be it in their L1 or in an L2.

From a learning perspective it can be assumed that L2 speakers perform better when material to be read is not only presented in its orthographical form but also provided in spoken form by a native speaker. In the latter case the learner's articulation should be faster and thus converging to the native-like articulation rate [8]. Reading only would then be slower than reading after listening to L1 speech.

In total we can thus identify four main sources of variation due to L2 proficiency, rate metrics, individuality and convergence. These sources of variation with expected patterns are illustrated schematically in Fig. 1 and described in more detail in the following subsections. Other sources of variation of articulation rate like for instance sex and age [9] are not considered here because they are not of primary relevance.
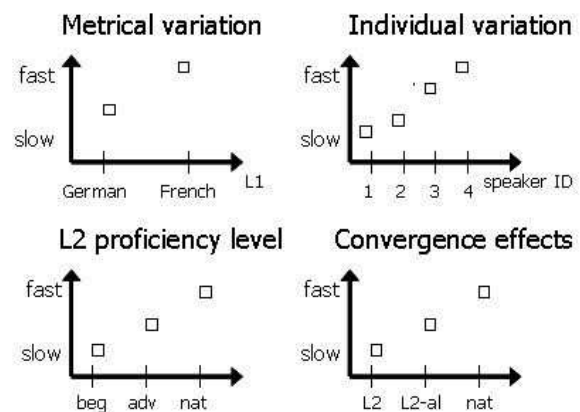

Figure 1: *Sources of variation of articulation rate with expected patterns (beg=L2 beginners, adv=advanced L2 learners, nat=native speakers, L2-al=L2 speech after listening).*

1.1. Speech rate metrics

When considering speech tempo or speech rate we distinguish between the *speaking rate* as the gross rate *in*cluding all pauses and *articulation rate* as the net rate of articulated speech, i.e. *ex*cluding all pauses.

In the concept of *fluency* (which is beyond the scope of this paper) articulation rate would be part of speed fluency which can be distinguished from repair fluency and breakdown fluency (cf. e.g. [10]).

Another issue in speech rate metrics is the choice of the basic linguistic unit. Possible candidates are the word, the syllable and the segment. The most popular unit seems to be the syllable [11, 9], thus having syllables per second or conversely the mean syllable duration as the preferred metrics. The advantage of the syllable in comparison to the word is that it is less variable in length within and across languages (compare for instance the mean length of words in Finnish or Turkish with those in French or English). The disadvantage of segments and syllables as basic units is that counting is not as easy as words. Another disadvantage of the segment is that they are more frequently elided than syllables or even words.

The aspect of elision leads to the next challenge selecting the optimal speech rate metric. Do we operate with phonological units which can be assumed as underlying representations and used as planned units, or should we focus on the actual realisation of these units (cf. [9, 12])? An argument in favour of the phonological units would be that the omitted segments or syllables or words are likely to be accounted for in speech planning and probably in speech perception, too. Thus, one strategy of speaking faster is to omit more units. An argument against the phonological units would be that speech with a moderate speed of articulation but with many omissions would reflect sloppiness but not necessarily fast articulation. For pragmatic reasons the use of phonological units often seems to be the preferred option, because for scripted material the number of units would be the same and the working load of analysing all speech samples with regard to omissions does not apply (apart from the fact that omission of segments is not always as clear as wished).

1.2. Research questions

1.2.1. Rate metrics

What is the most informative unit for rate metrics which allows comparisons between French and German? We are looking for a metric that facilitates the comparison of the speaking tempo between various speech modes such as L1 speech of different languages, L2 speech of different languages, different levels of proficiency in L2 speech, and further individual differences, in L1 *and* L2 speech. Probably there is not a single metric reflecting all these levels of variation.

However, we expect that syll/s is not the best metric to reflect articulation rate in L1 and L2 speech for the French-German language pair. Comparisons of syllable rates between German and French show substantially higher values for French, e.g. 7.3 syll/s (Fr.) vs. 5.6 syll/s (G.) in [13] or 7.18 syll/s (Fr.) vs. 5.97 syll/s (G.) in [14]. This phenomenon can be explained by a higher complexity of syllables in German where more time is required to articulate more segments in a syllable.

1.2.2. Native speech and the level of L2 proficiency

Regarding native speech and the level of L2 proficiency we ask: Do beginners articulate slower than advanced learners, and do native speakers articulate faster than advanced learners? For both questions we would expect an affirmative answer for both languages.

1.2.3. Individual habits of articulation rate

Are individual articulation rate habits in L1 visible in L2 speech? Following [15, 10], who found partial evidence for rate variation on the learner's personal speaking style irrespective of L2 proficiency, we can expect an affiramtive answer to this question, too.

1.2.4. Convergence to L1 speech

Can we observe any convergence of articulation rate to native speakers when learners get more extensive exposure to native speech, e.g. reading a sentence after listening to that sentence read by a L1 speaker? The use of more than one sensory channel is generally assumed to be helpful in teaching and learning foreign language phonetics [16, 17]. We would expect an affirmative answer to this question as well.

1.3. Outline

The remainder of the paper is structured as follows: in section 2 we describe the materials and subjects used for this study, followed by an analysis of various metrics of articulation rate. In section 3 we present the results with regard to rate metrics and the research questions, which will be discussed in section 4 which concludes the paper with an outlook on future work.

2. Method

2.1. Material

We used a part of a phonetic and phonological learner corpus for the language pairs French-German and German-French [18]. The data are from 7 speakers with French as their first language and 7 speakers with German as L1. Each language group consisted of five speakers at a beginner level (A1 or A2 according to the European reference frame CERF) and two at an advanced level (C1 or C2 at CERF). All speakers were recorded in their L2 *and* their L1.

Samples from two conditions were selected, viz. (i) read sentences ("read"), (ii) repeated sentences ("repeated"). In the latter condition only L2 speech was recorded: the sentences were prompted orthographically (as in the "read" condition) but synchronously with an audio file spoken by a native speaker. These native model speakers were not subjects in this corpus. The task was to repeat the written sentence just heard. Note that the repetition was produced a few seconds after listening in order to avoid a direct imitation.

For each of the 14 speakers we analysed 10 sentences in the "read L1", 10 in the "read L2", and 10 in the "repeated" condition in the respective L2. As a consequence the sentences "Read French" and the sentences "Read German" were produced by all 14 speakers, either in their L1 or in their L2. The sentences "Repeat French" were produced only by the German speakers, and the sentences "Repeat German" only by the French speakers. In total the analysis included 30 (sentences) x 14 (speakers) = 420 sentences.

2.2. Analysis

Start and end of the articulation of the sentences as well as silent and breathing pauses (if present) were first determined by means of an automatic speech recognition procedure using forced alignment. In a second step all labels were auditorily and visually checked with the speech editor Praat and manually corrected, if needed. We noted the appearance of possible addi-

tional syllables due to repetition, reflecting a disfluency that is not uncommon in L2 speech.

For each sentence the speaking rate (as a gross rate) and the articulation rate (as a net rate) were calculated for phones per second (phon/s), syllables per second (syll/s), and words per minute (wds/min).

The unit for syll/s was a *phonological* syllable, i.e. as produced in a canonical form. For phon/s the basic unit was the segment as predicted from the canonical form. The unit for wds/min was the word as counted in a text processor, e.g. the French "c'est" or "s'est" was counted as one word.

3. Results

3.1. Rate metrics

In about half of the sentences of L2 speech a pause was inserted. Pause insertion in L1 speech happened infrequently (less than 5 percent).

Fig. 3 shows that the syllable rates for French L1 and German L1 differ substantially whereas this is not the case for the phone rate (Fig. 2) and the word rate (Fig. 4).

The individual values of syll/s for the German speakers (S8-S14) show for the majority of speakers higher rates in their L2 than in their L1.

The distinction between L1 and L2 for each speaker group is achieved best for the metric phon/s. Thus, we present the data in phon/s in the following sections.
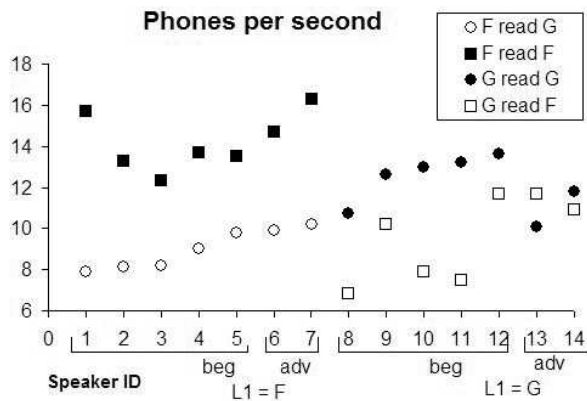

Figure 2: *Articulation rate of speakers in phon/s for the "read" condition. Speakers 1-7 with L1 = French (F), 8-14 with L1 = German (G). Speakers 6-7 and 13-14 are more advanced L2 speakers.*

3.2. L2 proficiency

In Fig. 2 the differences between native and non-native speech are clearly visible. For the French sentences the slowest L1 speaker is faster than the fastest L2 speaker which clearly marks a distinctive line between L1 French and L2 French speech. Similarly, in German the slowest L1 speaker (S13) articulates as fast as the fastest L2 speaker (S7). Both groups are clearly distinguishable.

For the level of L2 proficiency we see that the fastest L2 speakers (S7 for French, S13 for German) belong to the advanced learners. Analogously the slowest L2 speakers are L2 beginners (S1 for French, S8 for German). However, selected
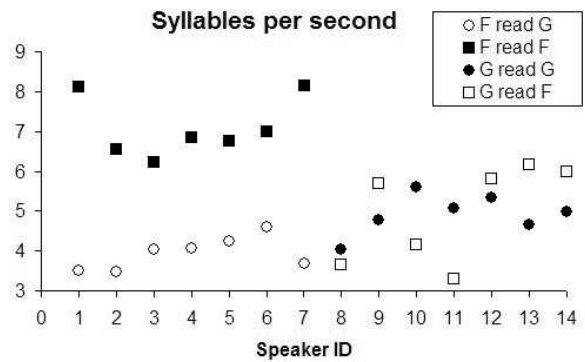

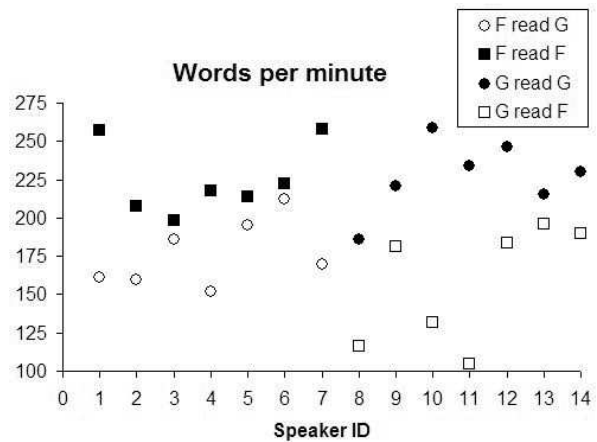Figure 3: *Articulation rate in syll/s for the "read" condition. Speaker grouping as in Fig. 2.*

Figure 4: *Articulation rate in wds/min for the "read" condition. Speaker grouping as in Fig. 2.*

beginners such as S5 (for French) and S12 (for German) are also very fast.

3.3. Individuality

Fig. 5 shows that the range of articulation rate for French L1 speech lies between 12 and 17 phon/s, whereas German L1 speech shows values between 10 and 14 phon/s. Thus, the difference between the slowest and the fastest speaker in each L1 group is between 4 phon/s (G.) and 5 phon/s (Fr.).

When looking at the L2 groups these differences change: the French speakers with L2 German articulate with a rate between 8 and 11 phon/s, thus reducing the range to 3 phon/s. The articulation rate of the German speakers with L2 French lie between 7 and 12 phon/s, thus expanding the range to 5 phon/s.

Individual tempo habits also lead to the rather unexpected outcome that one speaker (S13) shows 'faster' values in L2 than in L1 speech. This example and also S14 show that slow talkers in their L1 can be fast talkers in their L2. Likewise, fast L1 speakers are not necessarily fast L2 speakers: the fast L1 speaker S1 is in his L2 speech as slow as S3 who is the slowest L1 speaker in French. Counterexamples are the German speakers S8, who is slow in his L1 and L2, and S12 as a fast counterpart.

3.4. Convergence

To compare the "read" and the "repeat" conditions for the two L2 groups Fig. 5 shows that all French beginners articulate faster when they have additional auditive input of L1 speech. Among the German beginners of French only three out of five were faster in the "repeat" condition.

The convergence effect is diminished for the advanced learners where only one out of four speakers was faster than in the "read only" condition. In general it can be seen that the learners with the slowest rates in "read" speeded up in "repeat".

The mean rate for the sentences by the two German model speakers in the "repeat" condition was 11.9 phon/s (11.7 phon/s for the one speaker, 12.1 phon/s for the other speaker). The speakers S5 (beginner) and S7 (advanced) nearly approached this articulation rate after listening but not for reading only.

The French model speaker was a rather slow speaker with 10.3 phon/s. All L2 speakers were slower than him, although three speakers (S12-S14) were faster in the "read" condtion.
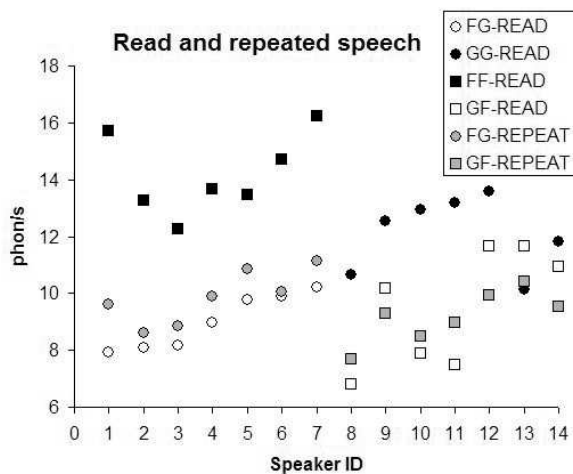

Figure 5: *Articulation rate for the "read" and "repeat" conditions. Speaker grouping as in Fig. 2.*

4. Discussion and conclusion

The rather low number of subjects per language and proficiency level allows only a limited interpretation of the results. Thus we need to be cautious with respect to generalisations. However, the explorative approach taken here has yielded interesting observations worth reporting and they represent good starting points for follow-up studies.

The best reflection of the L1-L2 difference for each speaker in our data seems to be phon/s. However, the phone rate for French is still considerably higher than for German (by 2.1 phon/s). A correction by this difference for the L1 German speech would result in a balanced picture with comparable ranges for L1 French and L1 German on the one hand, and comparable differences between L1 and L2 for the French speakers and the German speakers on the other hand. However, such a correction would ignore a correction for L2 German which would be a theoretical flaw. Although studies on speech rhythm often apply a normalisation of articulation rate between different languages (e.g. [13]), this usually concerns L1 speech and not L2 speech of the same speakers.

L2 proficiency is only to a limited extent reflected by articulation rate in our data. We deliberately left out learners at a medium level (B1 or B2 at CERF), but the distinction between beginners and advanced learners is not as clear-cut as expected.

The expectation regarding the transfer of the individual articulation rate habits from L1 to L2 is not entirely met. There is a huge variation among L2 speakers that *only partially* reflects L1 rate habits – in contrast to other studies [15, 10].

Our hypothesis about converging articulation rates of an L2 speaker after listening to an L1 speaker was confirmed. There is a positive effect of additional listening on speeding up for beginners with slow articulation rates. The extremely slow articulation rate of the French model speaker had a slowing-down effect on the fastest German speakers after listening. In both cases convergence in terms of articulation rate has evidently occurred. Since speaking slower may also indicate a more careful speaking style the slowed down L2 speakers possibly felt motivated to speak more clearly as well. Future studies are needed to clarify what kind of speaking style and what articulation rate can help which type of learner to improve the intelligibility and the fluency in L2 speech production.

Further topics for future research include the choice of rate metric. First, the metrics syll/s and phon/s are calculated based on the phonological structure rather than its actual realisation. Second, it is currently unclear how far the metrics applied here reflect the *perceptual* tempo [12, 19]. Another question in the context of speech perception is how comprehension is correlated with articulation rate. A last but not least point concerns the speech materials. In this study we used read sentences. It remains unclear how we can generalise from this sort of data to scripted and non-scripted styles, particularly when the discourse units are larger than a sentence. Nevertheless, read sentences, also with additional listening, are quite useful for exercises in learning and teaching environments.

The basic idea we started with was that articulation rate as the key component of speech tempo would be a good and easy-to-handle indicator of the level of the *spoken* language learning process. We have shown that for a comparison of different languages a correction for the variation caused by different phonological complexities should be considered.

As expected, more proficient L2 learners show remarkably high articulation rates, although not as high as L1 speakers. It is important to note that among the beginners there are speakers with similarly high articulation rates. One possible explanation for this finding is that different phonetic talents [20] also influence the speed of articulation.

An interesting observation is that there is generally an improvement in the temporal control of articulation when the L2 speakers had an additional auditory input of the sentence to be read. Although this is not the case for all L2 speakers, most of the *beginners* speeded up. It can be expected that this group benefit most from the input in multiple modalities when learning an L2. This point is important for the selection of exercises where speed of articulation should be part of a test and evaluation scheme which also includes computer-assisted pronunciation training.

5. Acknowledgements

6. References

[1] P. Lennon, "Investigating fluency in EFL: A quantitative approach," *Language Learning*, vol. 40, pp. 387–417, 1990.

[2] M. Munro and T. Derwing, "Modelling perceptions of the comprehensibility and accentedness of L2 speech: The role of speaking rate," *Studies in Second Language Acquisition*, vol. 23, pp. 451–468, 2001.

[3] H. Pürschel, *Pause und Kadenz. Interferenzerscheinungen bei der englischen Intonation deutscher Sprecher*. Tübingen: Max Niemeyer Verlag, 1975.

[4] M. Raupach, "Temporal variables in first and second language speech production," in *Temporal Variables in Speech: Studies in Honour of Frieda Goldman-Eisler*, H. Dechert and M. Raupach, Eds. The Hague: Mouton, 1980, pp. 263–270.

[5] U. Gut, *Non-native Prosody. A corpus-based analysis of the phonetic and phonological properties of L2 English and L2 German*. Frankfurt: Peter Lang, 2009.

[6] R. Wiese, *Psycholinguistische Aspekte der Sprachproduktion*. Hamburg: Buske, 1983.

[7] U. Gut, "Prosody in second language speech production: the role of the native language." *Fremdsprachen Lehren und Lernen*, vol. 32, pp. 133–152, 2003.

[8] R. L. Street and H. Giles, "Speech accommodation theory: A social cognitive approach to language and speech behavior," in *Social Cognition and Communication*, M. E. Roloff and C. R. Berger, Eds. Beverly Hills, CA: Sage, 1982, pp. 193–226.

[9] J. Trouvain, *Tempo Variation in Speech Production. Implications for Speech Synthesis*. PhD Dissertation, Saarland University, Saarbrücken, 2004.

[10] N. H. De Jong, R. Groenhout, R. Schoonen, and J. H. Hulstijn, "Second language fluency: speaking style or proficiency? Correcting measures of second language fluency for first language behavior," *Applied Psycholinguistics*, vol. 34, p. http://dx.doi.org/10.1017/S0142716413000210, 2013.

[11] T. H. Crystal and A. S. House, "Articulation rate and the duration of syllables and stress groups in connected speech," *Journal of the Acoustical Society of America*, vol. 88, pp. 101–112, 1990.

[12] J. Koreman, "Perceived speech rate: the effects of articulation rate and speaking style in spontaneous speech," *Journal of the Acoustical Society of America*, vol. 119, pp. 582–596, 2006.

[13] V. Dellwo and P. Wagner, "Relations between language rhythm and speech rate," in *Proc. ICPhS, Barcelona*, 2003, pp. 471–474.

[14] F. Pellegrino, C. Coupé, and E. Marsico, "A cross-language perspective on speech information rate," *Language*, vol. 87, pp. 539–558, 2011.

[15] T. M. Derwing, M. J. Munro, R. I. Thomson, and M. J. Rossiter, "The relationship between L1 fluency and L2 fluency development." *Studies in Second Language Acquisition*, vol. 31, pp. 533–557, 2009.

[16] U. Hirschfeld and J. Trouvain, "Teaching prosody in German as a foreign language," in *Non-Native Prosody. Phonetic Description and Teaching Practice.*, J. Trouvain and U. Gut, Eds. Berlin: Mouton De Gruyter, 2007, pp. 171–187.

[17] H. Mitterer and J. McQueen, " Foreign subtitles help but native-language subtitles harm foreign speech perception," *PLoS One*, vol. 4, pp. A146–A150, 2009.

[18] J. Trouvain, Y. Laprie, B. Möbius, B. Andreeva, A. Bonneau, V. Colotte, C. Fauth, D. Fohr, D. Jouvet, O. Mella, J. Jügler, and F. Zimmerer, "Designing a bilingual speech corpus for French and German language learners," in *Proc. Corpus et Outils en Linguistique, Langues et Parole: Statuts, Usages et Mésuages, Strasbourg*, 2013, pp. 32–34.

[19] H. R. Pfitzinger, "Local speech rate perception in german speech," in *Proc. ICPhS, San Francisco*, vol. 2, 1999, pp. 893–896.

[20] M. Jilka, "Talent and proficiency in language," in *Language Talent and Brain Activity.*, G. Dogil and S. Reiterer, Eds. Berlin: Mouton De Gruyter, 2009, pp. 1–16.