

# Do speakers converge rhythmically? A study on segmental timing properties of Grison and Zurich German before and after dialogical interactions

Elisa Pellegrino<sup>1</sup>, Sandra Schwab<sup>2</sup>, Volker Dellwo<sup>1</sup>

<sup>1</sup>Department of Computational Linguistics, University of Zurich, Zurich, Switzerland

<sup>2</sup>Institut de langue et de littérature française, Université de Berne, Bern, Switzerland  
elisa.pellegrino@uzh.ch ORCID: <https://orcid.org/0000-0001-5332-8411>  
sandra.schwab@unibe.ch ORCID: <https://orcid.org/0000-0003-4485-8335>  
volker.dellwo@uzh.ch ORCID: <https://orcid.org/0000-0002-8494-6025>

Submitted: 28/09/2021; Accepted: 11/10/2021; Published on line: 28/06/2022

**Citation / Cómo citar este artículo:** Elisa Pellegrino, Sandra Schwab, Volker Dellwo (2021), Do speakers converge rhythmically? A study on segmental timing properties of Grison and Zurich German before and after dialogical interactions. *Loquens*, 8(1-2), e078. <https://doi.org/10.3989/loquens.2021.e078>.

**ABSTRACT:** This paper reports on the results of a research investigating whether rhythmic features, in terms of segmental timing properties, are object of speaker's adjustments after the exposure to a conversational partner. In the context of dialects in contact, this is crucial to understand whether rhythmic attributes may bring about language variation and change. In the context of human-machine interactions, this can benefit the design of spoken dialogues systems to achieve human-likeness. To study rhythmic accommodation, we selected a corpus of pre- and post-dialogue recordings, performed by 18 speakers of Grison and Zurich German (henceforth GRG and ZHG), two Swiss German dialects characterised by noticeable segmental and supra-segmental differences. To quantify rhythmic convergence, we designed three measures based on the segmental timing differences between the two dialects. We compared the Euclidean distances in the three measures between GRG and ZHG speakers in a pair before and after two interactions. Results reveal that dyads members do not significantly shift the production of segmental timing features after the dialogues. Neither linguistic nor social factors can account for the observed accommodation pattern. Cross-dialectal segmental timing differences, captured by the three ratio measures, may be either robust against the influence of interlocutors' acoustic behaviour or too subtle to be perceived or retained after interactions.

**Keywords:** phonetic convergence, speech rhythm, dialects in contact, speech production.

**RESUMEN:** *Convergen rítmicamente los hablantes? Estudio sobre las propiedades segmentales temporales en el alemán de Zúrich y del cantón de Grisón antes y después de las interacciones dialógicas.*- Este artículo presenta los resultados de una investigación que examina si los rasgos rítmicos vinculados con las características temporales segmentales sufren adaptaciones por parte del locutor después de cierta exposición con un interlocutor. En un contexto de dialectos en contacto, es importante comprender la aportación del estudio de las propiedades rítmicas al campo de la variación y cambio de lengua. En un contexto de interacción persona-ordenador, esto puede beneficiar el diseño de sistemas de diálogos que logren un habla casi nativa. Para estudiar la acomodación rítmica, se seleccionó un corpus de grabaciones recopiladas antes y después de un diálogo entre 18 pares de locutores de alemán del cantón de Grisón y de alemán de Zúrich (de ahora en adelante, GRG y ZHG), dos dialectos suizo-alemanes que se caracterizan por sus diferencias segmentales y suprasegmentales. Para cuantificar la convergencia rítmica, se usaron tres medidas basadas en las diferencias temporales segmentales entre los dos dialectos. Se comparó la distancia euclidiana para las tres medidas entre locutores GRG y ZHG antes y después de dos interacciones. Los resultados revelan que los miembros del par no cambian significativamente la producción de sus características rítmicas temporales después del diálogo. Ningún factor lingüístico o social permite explicar el patrón de acomodación que se

observa. Las diferencias rítmicas temporales inter-dialectos, que se capturan con las tres medidas, son o demasiado robustas contra la influencia del comportamiento rítmico de los interlocutores, o demasiado sutiles para ser percibidas o recordadas después de las interacciones.

**Palabras clave:** convergencia fonética, ritmo, dialectos en contacto, producción del habla

## 1. INTRODUCTION

The way an individual speaks is highly idiosyncratic as it is largely determined by his/her anatomy, sex, age, language background, social status and health conditions (Dellwo et al., 2007).

During social interactions, however, the way individuals sound like they do is also influenced by the characteristics of the interlocutor (i.e., age, dialect, social status), the formality of the communicative setting (i.e., formal vs informal) and the quality of background conditions (i.e., noisy vs quiet) (Giles & Ogay, 2007). When we address to infants, relative to adults, for example, we typically speak slower, use longer pauses, exaggerate pitch variations and hyper-articulate vowels (see a.o., Fernald et al., 1989; Soderstrom, 2007). Most of these acoustic characteristics that are used to gain an infant's attention and to facilitate language acquisition, are also present when talking to elderly people (Kemper, 1994), and to some extent to second language speakers (Ferguson, 1975), or when the interaction takes place in a noisy environment (Hazan & Baker, 2011) to foster comprehension.

In addition to speech adjustments in response to interlocutors' characteristics, communicative and background conditions, there is evidence that interlocutors tend to adjust their verbal and non-verbal behaviour during and after exposure to a communication partner or a model talker. This phenomenon is known as accommodation (Giles & Ogay, 2007), alignment (Pickering & Garrod, 2006), entrainment (Brennan, 1996), synchrony (Edlund et al., 2009), mimicry (Pentland, 2008) and chameleon effect (Chartrand & Bargh, 1999). Evidence of mutual adjustments between speakers has been found in conversation and shadowing tasks (see a.o. Pardo et al., 2018) and encompasses many linguistic, para- and extralinguistic features. Indeed, accommodation has been found at the level of lexical choices (e.g., Bell, 2001; Ward & Litman, 2007), grammatical and syntactical structures (Branigan et al., 2000; Reitter et al. 2006), pronunciation (i.e., vowel quality, voice onset time, rate,  $f_0$ , intensity) (see a.o. Babel, 2012; Nielsen, 2011; Sancier & Fowler, 1997; Zellou, 2016; Levitan & Hirschberg, 2011; Manson et al., 2013; Ross et al., 2021), facial expressions (Lakin, 2013) and body movements (Dijksterhuis & Bargh, 2001). Accommodation is not a prerequisite of human-human interactions, as evidence of this phenomenon has been found in human-computer interactions (see a.o. Bell et al. 2003; Ravet et al., 2019; Gessinger et al. 2021) and in animal communication (Ruch et al., 2017 for a review).

For the domain of human-human communication, two major theoretical models have been proposed to ac-

count for interspeaker' adjustments: the social approach of the Communication Accommodation Theory (CAT) (e.g., Giles et al. 1991; Shepard et al., 2001) and the automatic account of the Interactive Alignment Model (IAM) proposed by Pickering & Garrod (2004). The former postulates that speakers express social closeness to or distance from their interlocutors, by respectively becoming acoustically more similar (convergence) or dissimilar (divergence) (Soliz & Giles, 2016). The latter, instead, assumes that convergence in conversation is regulated by a priming mechanism based on the automatic link between perception and production. Evidence in support of CAT can be found in studies showing that social factors, among which speakers' perceived friendliness, dominance, attractiveness, attitude or stereotypes towards a specific language variety (e.g., Babel et al., 2013; 2014, Schweitzer & Lewandowski, 2014, Michalsky & Schoormann, 2017; Gregory & Webster, 1996) affect the amount and direction of accommodation. IAM, instead, is supported by the line-up of studies documenting convergence in non-interactive settings (e.g., shadowing task) in which participants are not instructed to imitate the model talker or explicitly requested to avoid imitation (e.g., Goldinger, 1998; Shockley et al., 2004; Walker & Campbell-Kibler, 2015; cf. Dufour & Nguyen, 2013, for a comparison between imitation and shadowing tasks).

Studies on phonetic convergence, however, have pointed out the influence of factors other than social on speakers' accommodation behaviour. It has been, indeed, observed that individuals greatly vary in the amount and direction of convergence depending on the frequency characteristics of the lexical items (Goldinger, 1998; Goldinger & Azuma, 2004; Nielsen, 2011), previous exposure to lexical items (Goldinger, 1998; Goldinger & Azuma, 2004), cognitive load involved in a task (Abel & Babel, 2017), and phonetic distance between interlocutors' language repertoires (Babel, 2012; Walker & Campbell-Kibler, 2015; Walters et al., 2013). The effect of linguistic and phonetic factors was not accounted for by either IAM or CAT. A more dominant view that reconciles the social and the automatic perspectives and integrates the effect of linguistic-phonetic factors on accommodation is the so-called hybrid approach (Babel, 2012; Pardo, 2012; Pardo et al., 2017). In this view, social, linguistic and phonetic factors are seen as catalysts or inhibitors of convergence in that they can boost or diminish the strength of the link between perception and production.

The aim of the present paper is to contribute to advancing the understanding of forms and factors evoking convergence, shifting the attention from the typical acoustic correlates of phonetic convergence (i.e., vowel

quality, rate, pitch, intensity, voice onset time) to speech rhythm, conceptualized here as the variability of segmental durational characteristics. Rhythmic convergence is studied using a pre-existent dataset designed to study cross-dialectal vowel convergence (cf. 2.1.). This will ultimately permit to compare the accommodation behavior of the same speakers across different measures, and test which type of factors between linguistic (cross-dialectal phonetic distance) and social (dialect markedness) will be driving convergence or divergence.

### 1.1. Rhythmic Accommodation

Three basic questions may arise when studying rhythmic accommodation: (a) Can speech rhythm in terms of segmental timing properties be object of adjustments between speakers? (b) In which communicative contexts is it possible to study rhythmic accommodation? and (c) Why is the research on accommodation in segmental timing a worthwhile pursuit?

With respect to (a) speech rhythm research has provided evidence that the durational characteristics of consonantal and vocalic intervals, as well as amplitude envelope characteristics, vary in response to the interlocutor's age and cognitive development. For example, studies on the rhythmic characteristics of infant-compared to adult-directed speech have shown that: a) English, Catalan and Spanish mothers present less durational variability of consonantal and vocalic intervals as well as longer vowel duration when speaking to their children compared to addressing adults (Payne et al., 2009); b) in Australian English delta modulations corresponding to the prosodic stress is greater in infant- than in adult-directed speech, while theta modulations, tracking syllable patterns, dominated the adult-directed speech modulation spectrum (Leong et al., 2017). Not only do speech rhythm vary depending on the interlocutors' characteristics, but the presence itself of an interlocutor (i.e., reading partner) has been shown to influence the degree of rhythm entrainment in synchronous reading tasks (Cerdeña-Oñate et al., 2021). In light of these findings, it seems plausible to assume that speakers can also mutually adapt the production of segmental timing features after exposure to a dialogue partner. On the other hand, in view of evidence showing that the timing properties of different speech intervals (e.g. consonants, vowels, voicing) are resistant to different sources of within speaker variability (speaking style, prosodic and linguistic factors) (Dellwo et al., 2015; Leeman et al. 2014), we cannot exclude that the speakers may maintain their segmental durational characteristics in post-dialogue productions. We will test precisely these two competing hypotheses in the present study.

With respect to (b), one of the contexts in which the study of rhythmic accommodation is possible is that of dialects in contact. In this setting, one might examine whether speakers of dialects that are mutually intelligible but present distinct rhythmic features

converge rhythmically after being exposed to each other's dialect. In this respect, the linguistic situation of German-speaking Switzerland is an excellent testing ground for studying cross-dialectal rhythmic accommodation. Swiss German dialects, indeed, do not only differ for segmental features, speech rate and intonation contours (see Leeman, 2012 for a review), but also for their rhythmic properties. It has been documented that Midland vs Alpine dialects as well as Eastern vs Western dialects can be grouped according to their rhythmic characteristics, measured acoustically in terms of the timing variability of consonantal and vocalic intervals (Leeman et al., 2012).

With respect to (c), it has been argued that assessments of phonetic convergence based on a single (supra)segmental feature hardly capture the complexity of the phenomenon (Pardo et al., 2017). Nevertheless, choosing one acoustic attribute over another is still a valid approach when the comprehension of dynamics of sound variation and change is at stake (Pardo et al., 2017), or when decisions must be taken about which aspects of human-human interaction have to be modelled in speech interactive systems to achieve human-likeness (Beňuš, 2014). Understanding whether rhythmic properties in terms of segmental durational characteristics are object of mutual adaptations can be also crucial for the interpretation of evidence in forensic phonetic speaker comparisons. Any acoustic adjustments between interlocutors might lead to mistake within-between-speaker variability and produce higher error in recognition rate.

## 2. THE STUDY

### 2.1. Material

To study rhythmic accommodation in a dialect contact situation, we used a corpus of speech material in Zurich and Grison German (henceforth ZHG and GRG), two Swiss German dialects exhibiting crucial segmental and suprasegmental differences (cf. 2.2.) that legitimate the assumption of interspeaker adjustments after exposure to the interlocutor's dialect.

The corpus was designed, collected and annotated by Hanna Ruch to study vowel accommodation between GRG and ZHG (Ruch, 2015). It included speech samples of:

- 2 audio-recorded diapix tasks (i.e., speakers comparing pictures that contain a certain number of differences, cf. Van Engen et al., 2010) performed by 18 pairs of previously unacquainted GRG and ZHG female speakers.
- 18 pre- and 18 post-dialogue recordings (picture naming task and retelling a story based on a comic), performed individually by GRG and ZHG participants.

The diapix tasks were designed to elicit the target words present in picture naming task and story retelling. All tasks were carried out in one single recording session.

## 2.2. Cross-dialectal phonetic differences

Grison and Zurich German present noticeable differences at several linguistic levels (Eckhardt, 1991; Fleischer & Smith, 2006; Christen et al., 2010; Leeman, 2012). Phonetically, these have to do with the quality of front vowels, realization of word-initial and post-vocalic k, speech rate and intonation contours. It is of interest - for the purpose of this study - that GRG and ZHG also exhibit segmental durational differences that lead to a distinct rhythmic organisation of the two dialects. As reported in the literature on acoustic differences between GRG and ZHG (see a.o. Ruch, 2018), these differences concern: a) intervocalic sonorants gemination (henceforth ISG) in words ending in -e; b) open syllable lengthening (henceforth OSL); c) vowel reduction in word final position (henceforth RedVow).

Given that segmental timing properties are among the acoustic correlates of speech rhythm, in this paper we will refer to the three cross-dialectal differences in ISG, OSL and RedVow as rhythmic differences. Regarding ISG, GRG intervocalic sonorants can be realized either as geminates or as single consonants, while ZHG allows only the singleton realisation. As for OSL, in GRG open syllables can be either lengthened or not, while in ZHG the lengthening tendency has not been documented. With respect to RedVow, in GRG vowels in word final position are not reduced in quality, and presumably either in duration, while in ZHG word final vowels are always reduced. (Cf. Table 1 for examples of cross-dialectal realizations of ISG, OSL and GR).

**Table 1:** Examples of items in GRG and ZHG for the three dialectal features (adapted from Ruch, 2018)

Feature	Example	GRG realization	ZHG realization
ISG	Sonne 'sun'	nn ['sunnə] n [sunə]	n ['sunə]
OSL	Sohle 'sole'	V: ['so:lə] V ['solə]	V ['solə]
Red Vow	Suppe 'soup'	ɐ ['suppə]	ə ['suppə]

Evidence in support that the differences in the quality of final vowels come also with distinct timing patterns has been provided in Leeman et al. (2012). Here it was shown that the durational variability of vocalic intervals was higher in Midland dialects (to which ZHG belongs to) than in the group of Alpine dialects (to which GRG belongs to), and this was interpreted in view of the tendency of Alpine dialects to retain full vowels in unstressed position.

## 2.3. Method

To understand whether pairs of GRG and ZHG speakers produce the rhythmic features more similarly after participating in the diapix tasks, the following steps were taken:

- From the pre- and post-dialogue recordings of individual speakers, we extracted the lexical items instantiating the three target rhythmic features (ISG, OSL and RedVow)<sup>1</sup>.
- For every item, we measured the duration of individual segments. The raw measures of segment duration served as a basis for the calculation of three ratio measures designed *ad hoc* to capture inter-dialectal differences in ISG, OSL and RedVow.
  - For ISG, we calculated the ratio between the duration of intervocalic sonorants (l, n) in -CCe words (e.g., Sonne, Welle) and that of the corresponding sonorant in -Ce words (l or n from the item Melone).
  - For OSL, we calculated the ratio between the duration of stressed vowels in open syllables and that of unstressed vowels within the same item.
  - For RedVow, we calculated the ratio between the duration of stressed vowels in open and closed syllables and that of unstressed vowel within the same item.

To determine whether pairs GRG and ZHG speakers converge, diverge or maintain their rhythmic behaviour after the interaction, we calculated:

- the Euclidean distance within individual pairs in the three ratio measures in pre- and post-dialogue recordings (i.e., dist 1 = GRG pre - ZHG pre; dist 2 = GRG post - ZHG post);
- the difference in distance between the two speakers' production of a word before the dialogues (i.e., dist 1 = GRG pre - ZHG pre) and after the dialogues (i.e., dist 2 = GRG post - ZHG post). Accommodation within a pair (*DDpair*) was calculated as follow:  $DDpair = dist 2 - dist 1$ . A negative difference in distance is evidence of convergence. A positive value indicates divergence. A value 0 demonstrates maintenance.

## 2.4. Data analysis and statistics

The present study reports on the data extracted from the picture naming task. In view of evidence showing the influence of linguistic factors on accommodation (Goldinger, 1998; Goldinger & Azuma, 2004; Niels-

<sup>1</sup> As mentioned above, in this paper speech rhythm is conceived in a narrow sense, namely as the variability in segmental durational characteristics at the word level. For this reason, the analysis of convergence is focused on the three cross-dialectal segmental timing features (ISG, OSL and RedVow). However, as pointed out by one of the reviewers, the study of rhythmic convergence in broad sense would entail measuring more general parameters, like the classic rhythm metrics.

en, 2011), analysing the data from picture naming tasks (henceforth PNT) has given the main advantage of controlling for the effect of the item variability in the assessment of:

- cross-dialectal differences before the interactions;
- differences in distance between ZHG and GRG speakers in a pair before and after the interaction.

The lexical items used in this study and the dialectal features they instantiate are listed in Table 2 in the Standard German spelling.

**Table 2:** List of items from picture naming task

Feature	Examples
ISG (5 items per speaker)	Brunnen, Pfanne, Sonne, Spinne, Welle (5 items per speaker)
OSL (8 items per speaker)	Besen, Esel, Graben, Käfer, Lupe, Nase, Schlafen, Melone
RedVow (15 items per speaker)	Besen, Brunnen, Flosse, Graben, Lampe, Lunge, Lupe, Melone, Nase, Pfanne, Schlafen, Sonne, Spinne, Suppe, Welle

To test (a), i.e., whether pairs of GRG and ZHG speakers realised the three durational contrasts differently before the interaction, and thus to make sure that there was room for rhythmic accommodation, we ran three separate Linear Mixed Effects Models with the ratio measures (ISG, OSL and RedVow) as dependent variables, dialect (ZHG and GRG) as fixed factor, and speaker and lexical item as random effect (i.e. random intercepts).

In the light of segmental durational differences between GRG and ZHG mentioned above, we make the following hypotheses regarding the rhythmic behaviour of ZHG and GRG speakers before the interaction:

- ISG contrast is higher in GRG than in ZHG, given that in GRG intervocalic sonorants can be pronounced also as geminates, while in ZHG only as a single consonant.
- OSL contrast is higher in GRG than in ZHG, given that in GRG open syllables can be lengthened, while in ZHG typically are not.
- RedVow is higher in ZHG than in GRG given that in ZHG word final vowels are reduced, while in GRG they are pronounced as full vowels.

To test (b), i.e., whether pairs of GRG and ZHG speakers produce the three rhythmic features more similarly after the diapix tasks, we compared the Euclidean distances within pair in ISG, OSL and RedVow

before and after the interactions. We ran three separate Linear Mixed Effect Models, with Euclidean distance in ISG, OSL and RedVow as dependent variables and Session (1 = before interaction; 2 = after the interaction) as fixed factor. Given that Euclidean distance between pairs may vary before and after the interaction, in the structure of the random effect we first included the random slope of Pairs by Session. However, this model was too complex to be supported by the data. For this reason, we simplified the random effects by including the intercept for the interaction between Session and Pair, instead of the random slope. The random part of the model comprised also the intercept for Item.

We hypothesise that if rhythmic features are object of accommodation, dyads members adjust their rhythmic behaviour such that the Euclidean distance in ISG, OSL and RedVow will be lower after than before the interaction. In view of findings showing the effect that speakers converge more for features that differ mostly between dialects (MacLeod, 2012; Ruch, 2015; Walker & Campbell-Kibler, 2015; Clopper & Dossey, 2020) and between the speakers and the model talkers (Babel, 2012), we hypothesise that more accommodation is evoked by RedVow than ISG and OSL. RedVow, indeed, is one of the features that best distinguishes the two dialects. ZHG indeed exhibits open syllable lengthening - though in articulatory contexts other than GRG - and presents longer nasal duration in -CCer words. However, given that the realisation of reduced vowels is also a strong dialect marker for ZHG (Ruch, 2018), in view of evidence about little convergence for features that are dialect markers (Babel, 2010), we cannot exclude that the speakers may diverge or maintain their original behaviour for RedVow.

To test these hypotheses, we ran one Linear Mixed Effects Model with *DDpair* as dependent variable and Ratio Type (ISG, OSL and RedVow) as fixed factor. The random part of the model comprised the intercept for the interaction between Pair and Ratio, as well as the intercept for Item. Statistical analyses were performed with RStudio (2009-2019) Version 1.2.1335.

## 2.5. Results and Discussion

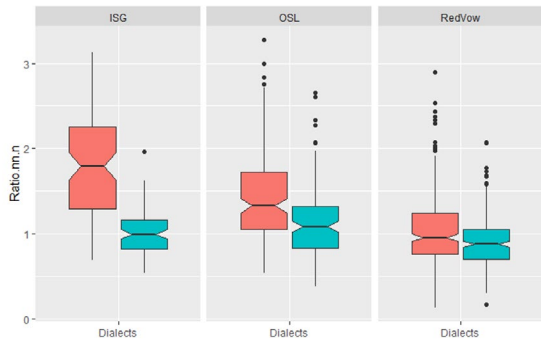
Regarding (a), i.e., cross-dialectal differences in ISG, OSL and RedVow before the interaction, the results from pre-dialogue recordings show a significant main effect of Dialect for the three measures (Table 3).

**Table 3:** Summary of statistics for the effect of Dialect on the three Ratio measures. Reference category is GR.

Ratio	Estimate	SE	t	p
ISG	-0.78	0.10	-7.81	<0.001
OSL	-0.31	0.06	-4.87	<0.001
RedVow	-0.12	0.03	-3.16	<0.01

As shown in Fig. 1, the scores obtained by GRG speakers in the three ratio measures are higher than ZHG speakers.

**Figure 1:** Cross dialectal differences in ISG (left) and OSL (centre), RedVow (right) in pre-dialogue PNT.



If the results for ISG and OSL are in line with predictions, what is more surprising is that RedVow is lower in ZHG than in GRG. One plausible explanation for this finding might be that in picture naming task, for which speakers were asked to pronounce words in isolation, ZHG speakers do not drastically reduce the duration of unstressed vowels in word final position, as these vowels are subjected to pre-pausal lengthening. In other words, in ZHG the durational difference between stressed and unstressed vowels in final word position is not that big as one might expect.

With respect to (b), i.e., the accommodation behaviour in ISG, OSL and RedVow, the results of statistical analysis reveal no significant main effect of Session (pre- and post-dialogue recordings) in the Euclidean distances (Table 4).

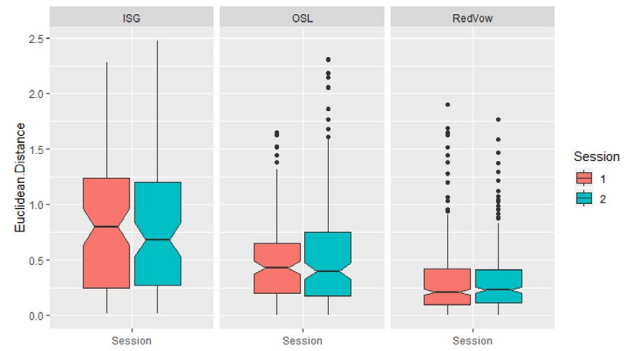
**Table 4:** Summary of statistics for the effect of Session on Euclidean distance. Reference category is Session 1.

Ratio	Estimate	SE	t	p
ISG	-0.02	0.10	-0.19	0.84
OSL	0.05	0.06	0.83	0.41
RedVow	0.008	0.03	0.25	0.82

In other words, the Euclidean distance between dyads members did not change significantly before and after the interactions (Fig. 2).

With respect to the hypothesis that RedVow is more prone to convergence compared to OSL and ISG, against the predictions, no significant differences in degree and

**Figure 2:** Euclidean distance within pair across sessions for ISG (left), OSL (centre), RedVow (right).



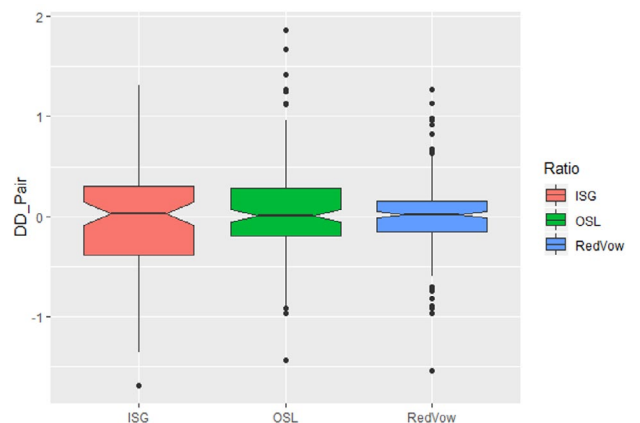
direction of accommodation (*DDpair*) were found between the three ratio measures (Table 5).

**Table 5:** Summary of statistics for the effect of Ratio Type on *DDpair*. Reference category is Ratio ISG.

	Estimate	SE	t	p
RatioOSL	0.08	0.07	1.15	0.25
RatioRedVow	0.04	0.06	0.71	0.47

Unlike findings on vowel accommodation between GRG and ZHG or between other dialects, showing more convergence for phonetically more distant features (Ruch, 2015; MacLeod, 2012; Walker & Campbell-Kibler, 2015; Clopper & Dossey, 2020), and more divergence for acoustic attributes perceived as strong dialect markers (Babel, 2010; Clopper & Dossey, 2020), in the case of ISG, OSL and RedVow, interpretations of accommodation based on phonetic distance or degree of dialect markedness do not seem tenable (Fig 3).

**Figure 3:** Difference in distance (*DDpair*) across the three ratio measures.





As shown in Fig. 3, RedVow, indeed, was neither more nor less prone to accommodation than OSL and ISG. Conversely, the values of the three measures circle around zero pointing in favour of rhythmic maintenance.

There could be at least two possible explanations for this result: (1) likewise the rhythmic metrics analysed in previous research (e.g., Leeman et al., 2014; Dellwo et al., 2015), the three timing measures examined here may be robust against source of within-speaker variability. The exposure to the distinct rhythmic behaviour of the dialogue partner might have not altered the post-dialogue realization of ISG, OSL and RedVow, as instead was observed for vowel formants. We cannot exclude, however, that accommodation in segmental timing properties has happened in the more spontaneous tasks of the corpus which has not been object of the present investigation. For future research, it will be interesting to examine whether the same pattern would replicate when rhythm is examined at the utterance level, using the metrics which have been typically employed in speech rhythm research (see a.o. Ramus, Nespore and Mehler, 1999; Grabe & Low, 2002; Dellwo, 2006; White and Mattys, 2007; He & Dellwo, 2016). (2) Another possible explanation may have to do, instead, with the perceptual salience of the cross-dialectal features, captured by the three rhythmic measures. Given that differences must be perceptible in order to be imitated (Mitterer & Müssler, 2013), the interspeakers' differences in ISG, OSL and RedVow may probably be too subtle to be perceived or retained after the interaction. This would be also in line with findings from Swiss German dialects recognition research that shows that listeners pay attention to segmental features to a higher degree than rhythmic and prosodic features when recognizing the dialectal origin of the speakers (see Leemann, et al., 2018; for varieties of English see a.o., Fuchs, 2015).

The differences in accommodation behaviour of the same ZHG and GRG speakers across segmental and rhythmic measures confirm the complexity and multi-facetedness of vocal accommodation. As pointed out by Sanker (2015) and Cohen Priva and Sanker (2018), patterns of convergence in one measure within a pair or within a speaker cannot be taken to be representative of pairs and speakers' overall convergence patterns in other measures.

### 3. CONCLUSIONS

Based on a corpus of pre- and post-dialogue picture naming task performed by 18 speakers of GRG and ZHG, results reveal that members of pairs, who show significant durational differences before the interaction, do not shift noticeably the production of ISG, OSL and RedVow after being exposed to the interlocutors' dialect. Although the evidence from rhythmic variability in child- and adult-directed speech, as well from synchronous reading, supports the view that rhythmic features can be object of interspeaker variations, these adjustments can be unidirectional and irrespective of the rhythmic behaviour of the dialogue partners

### 4. ACKNOWLEDGMENTS

This research was supported by the University Research Priority Program (URPP) "Language and Space" of the University of Zurich (UZH). We thank Hanna Ruch who designed, collected and annotated made available the corpus used in this study.

### 5. REFERENCES

- Abel J. & Babel M. (2017). Cognitive load reduces perceived linguistic convergence between dyads. *Language and Speech*, 60(3), 479-502. <http://dx.doi.org/10.1177/0023830916665652>
- Babel, M. (2010). Dialect divergence and convergence in New Zealand English. *Language in Society*, 39(4), 437-56. <http://dx.doi.org/10.1017/S0047404510000400>
- Babel, M. (2012). Evidence for phonetic and social selectivity in spontaneous phonetic imitation. *Journal of Phonetics*, 40(1), 177-189.
- Babel, M., McAuliffe, M., & Haber, G. (2013). Can mergers-in-progress be unmerged in speech accommodation? *Frontiers in Psychology*, 4(653), 1-14. <http://dx.doi.org/10.3389/fpsyg.2013.00653>
- Babel, M., McGuire, G., Walters, S. & Nicholls, A. (2014). Novelty and social preference in phonetic accommodation. *Laboratory Phonology*, 5(1), 123-150. <http://dx.doi.org/10.1515/lp-2014-0006>
- Bell, A. (2001). Back in style: Reworking audience design. In P. Eckert, & J. R. Rickford (Eds.), *Style and Sociolinguistic Variation* (pp. 139-169). Cambridge: Cambridge University Press.
- Bell, L., Gustafson, J., & Heldner, M. (2003). Prosodic adaptation in human-computer interaction. *International Congress of Phonetic Sciences*, (ICPhS), Barcelona, 2003, 2453-2456.
- Beňuš, Š. (2014). Social aspects of entrainment in spoken interaction. *Cognition Computing*, 6, 802-813. <http://dx.doi.org/10.1007/s12559-014-9261-4>
- Branigan, H. P., Pickering, M. J., & Cleland, A. A. (2000). Syntactic co-ordination in dialogue. *Cognition*, 75(2), B13-B25. [http://dx.doi.org/10.1016/S0010-0277\(99\)00081-5](http://dx.doi.org/10.1016/S0010-0277(99)00081-5)
- Brennan, S. E. (1996). Lexical entrainment in spontaneous dialog. *Proceedings of the International Symposium on Spoken Dialogue*, Philadelphia, PA, 41-44.
- Cerda-Oñate, K., Toledo Vega, G., & Ordin, M. (2021). Speech rhythm convergence in a dyadic reading task. *Speech Communication*, 131, 1-12. <http://dx.doi.org/10.1016/j.specom.2021.04.003>
- Clopper, C. G. & Dossey, E. (2020). Phonetic convergence to Southern American English: Acoustics and perception. *The Journal of the Acoustical Society of America*, 147(1), 671-671. <http://dx.doi.org/10.1121/10.0000555>
- Cohen Priva, U. & Sanker, C. (2018). Distinct behaviors in convergence across measures. *Annual Conference of the Cognitive Science Society*, Madison, WI, 1518-1523.
- Chartrand, T. L. & Bargh, J. A. (1999). The chameleon effect: The perception-behavior link and social interaction. *Journal of Personality and Social Psychology*, 76(6), 893. <http://dx.doi.org/10.1037/0022-3514.76.6.893>
- Christen, H., Glaser, E., & Friedli, M. (2010). *Kleiner Sprachatlas der deutschen Schweiz*. Frauenfeld: Huber Frauenfeld.
- Dellwo, V., Huckvale, M., & Ashby, M. (2007). How is individuality expressed in voice? An introduction to speech production and description for speaker classification. In C. Müller (Ed.), *Speaker Classification I* (pp. 1-20), LNAI 4343. Berlin-Heidelberg: Springer-Verlag.
- Dellwo, V. (2006). Rhythm and speech rate: A variation coefficient for C. In P. Kamowski & I. Sziget (Eds.), *Language and Language-Processing* (pp. 231-241). Frankfurt am Main: Peter Lang.
- Dellwo, V., Leemann, A., & Kolly, M.-J. (2015). Rhythmic variability between speakers: Articulatory, prosodic, and linguistic factors. *Journal of the Acoustical Society of America*, 137(3), 1513-1528. <http://dx.doi.org/10.1121/1.4906837>
- Dijksterhuis, A. & Bargh, J. A. (2001). The perception-behavior expressway: Automatic effects of social perception on social behavior. In M. Zanna (Ed.), *Advances in Experimental Social Psychology*, vol. 33, (pp. 1-40). San Diego: Academic Press.
- Dufour, S. & Nguyen, N. (2013). How much imitation is there in a shadowing task? *Frontiers in Psychology*, 4. <https://doi.org/10.3389/fpsyg.2013.00346>

- Eckhardt, O. (1991). *Die Mundart der Stadt Chur*. Zürich: Phonogrammarchiv der Universität 624, Zürich.
- Edlund, J., Heldner, M., & Hirschberg, J. (2009). Pause and gap length in face-to-face interaction. *10th Annual Conference of the International Speech Communication Association*, 2779–2782. <http://dx.doi.org/10.21437/Interspeech.2009-710>
- Ferguson, C. A. (1975). Towards a characterization of English foreigner talk. *Anthropological Linguistics*, 17, 1–14.
- Fernald A., Taeschner T., Dunn J., Papousek M., de Boysson-Bardies B., & Fukui I. (1989). A cross-language study of prosodic modifications in mothers' and fathers' speech to preverbal infants. *Journal of Child Language*, 16(3), 477–501. <http://dx.doi.org/10.1017/S0305000900010679>
- Fleischer, J. & Schmid, S. (2006). Zurich German. *Journal of the International Phonetics Association*, 36, 243–253. <http://dx.doi.org/10.1017/S0025100306002441>
- Fuchs, R. (2015). You're not from around here, are you? Dialect discrimination experiment with speakers of British and Indian English. In E. Delais-Roussarie, M. Avanzi, & S. Herment (Eds.), *Prosody and Language in Contact* (pp. 123–148). Berlin: Springer.
- Gessinger, I., Möbius, B., Le Maguer, S., Raveh, E., & Steiner, I. (2021). Phonetic accommodation in interaction with a virtual language learning tutor: A Wizard-of-Oz study. *Journal of Phonetics*, 86, 101029. <http://dx.doi.org/10.1016/j.wocn.2021.101029>
- Giles, H. & Ogay, T. (2007). Communication accommodation theory. In B. B. Whaley & W. Samter (Eds.), *Explaining Communication: Contemporary Theories and Exemplars* (pp. 293–310). Mahwah NJ: Lawrence Erlbaum.
- Giles, H., Coupland, N. & Coupland, J. (1991). Accommodation theory: Communication, context, and consequence. In H. Giles, J. Coupland, & N. Coupland (Eds.), *Contexts of Accommodation: Developments in Applied Sociolinguistics* (pp. 1–68). Cambridge: Cambridge University Press.
- Goldinger, S. D. (1998). Echoes of echoes? An episodic theory of lexical access. *Psychological Review*, 105(2), 251–279. <http://dx.doi.org/10.1037/0033-295X.105.2.251>
- Goldinger, S. D. & Azuma, T. (2004). Episodic memory reflected in printed word naming. *Psychonomic Bulletin & Review*, 11(4), 716–722. <http://dx.doi.org/10.3758/BF03196625>
- Grabe, E. & Low, E. L. (2002). Durational variability in speech and the rhythm class hypothesis. In N. Warner & C. Gussenhoven (Eds.), *Papers in Laboratory Phonology 7* (pp. 515–546). Berlin: Mouton de Gruyter. <http://dx.doi.org/10.1515/9783110197105.515>
- Gregory, S. W. & Webster, S. (1996). A nonverbal signal in voices of interview partners effectively predicts communication accommodation and social status perceptions. *Journal of Personality and Social Psychology*, 70(6), 1231–1240. [10.1037/0022-3514.70.6.1231](https://doi.org/10.1037/0022-3514.70.6.1231)
- Hazan, V. & Baker, R. (2011). Acoustic-phonetic characteristics of speech produced with communicative intent to counter adverse listening conditions. *The Journal of the Acoustical Society of America*, 130, 2139–2152. <http://dx.doi.org/10.1121/1.3623753>
- He, L., & Dellwo, V. (2016). The role of syllable intensity in between-speaker rhythmic variability. *International Journal of Speech, Language, and the Law*, 23(2), 243–275. <http://dx.doi.org/10.1558/ijssl.v23i2.30345>
- Kemper, S. (1994). Speech accommodations to older adults. *Aging and Cognition*, 1, 17–28.
- Lakin, J. L. (2013) Behavioral mimicry and interpersonal synchrony. In J. A. Hall & M. L. Knapp (Eds.), *Nonverbal Communication* (pp. 539–576). Berlin: De Gruyter Mouton. <http://dx.doi.org/10.1515/9783110238150.539>
- Leemann, A. (2012). *Swiss German Intonation Patterns*. Amsterdam, Philadelphia: John Benjamins Publishing Company.
- Leemann, A., Dellwo, V., Kolly, M. J., & Schmid, S. (2012). Rhythmic variability in Swiss German dialects. *6th International Conference on Speech Prosody*, Shanghai, China, 607–610.
- Leemann, A., Kolly, M.-J., & Dellwo, V. (2014). Speaker-individuality in suprasegmental temporal features: Implications for forensic voice comparison. *Forensic Science International*, 238, 59–67. <http://dx.doi.org/10.1016/j.forsciint.2014.02.019>
- Leemann, A., Kolly, M.-J., Nolan, F., & Y. Li (2018). The role of segments and prosody in the identification of a speaker's dialect. *Journal of Phonetics*, 68, 69–84. <http://dx.doi.org/10.1016/j.wocn.2018.02.001>
- Leong V., Kalashnikova, M., Burnham, D., & Goswami, U. (2017). The Temporal Modulation Structure of Infant-Directed Speech. *Open Mind: Discoveries in Cognitive Science*, 1, 78–90.
- Levitan, R. & Hirschberg, J. B. (2011). Measuring acoustic-prosodic entrainment with respect to multiple levels and dimensions. In P. Cosi, R. De Mori, G. Di Fabbrizio, & R. Pieraccini (Eds.), *Interspeech 2011*, 3081–3084. <http://dx.doi.org/10.21437/Interspeech.2011-771>
- MacLeod, B. (2012). *The Effect of Perceptual Salience on Phonetic Accommodation in Cross-Dialectal Conversation in Spanish*. Dissertation. Toronto: University of Toronto.
- Manson, J. H., Bryant, G. A., Gervais, M. M., & Kline, M. A. (2013). Convergence of speech rate in conversation predicts cooperation. *Evolution and Human Behavior*, 34(6), 419–426. <http://dx.doi.org/10.1016/j.evolhumbehav.2013.08.001>
- Michalsky, J., Schoormann H. (2017). Pitch convergence as an effect of perceived attractiveness and likability. *Interspeech*. Stockholm, 2253–2256.
- Mitterer, H. & Müsseler, J. (2013). Regional accent variation in the shadowing task: Evidence for a loose perception-action coupling in speech. *Attention, Perception and Psychophysics*, 75, 557–575. <http://dx.doi.org/10.3758/s13414-012-0407-8>
- Nielsen, K. (2011). Specificity and abstractness of VOT imitation. *Journal of Phonetics*, 39(2), 132–142. <http://dx.doi.org/10.1016/j.wocn.2010.12.007>
- Pardo, J. S., Gibbons, R., Suppes, A., & Krauss, R. M. (2012). Phonetic convergence in college roommates. *Journal of Phonetics*, 40(1), 190–197. <http://dx.doi.org/10.1016/j.wocn.2011.10.001>
- Pardo, J. S., Urmanche, A., Wilman, S., & Wiener, J. (2017). Phonetic convergence across multiple measures and model talker. *Attention, Perception, & Psychophysics*, 79(2), 637–659. <http://dx.doi.org/10.3758/s13414-016-1226-0>
- Pardo, J. S., Urmanche, A., Wilman, S., Wiener, J., Mason, N., Francis, K., & Ward, M. (2018). A comparison of phonetic convergence in conversational interaction and speech shadowing. *Journal of Phonetics*, 69, 1–11. <http://dx.doi.org/10.1016/j.wocn.2018.04.001>
- Payne, E., Post, B., Astruc, L., Prieto, P., & Vanrell, M. (2009). Rhythmic modification in child directed speech. *Oxford University Working Papers in Linguistics, Philology & Phonetics*, 12, 123–144.
- Pentland, A. (2008). *Honest Signal: How They Shape Our World*. Cambridge, MA: MIT Press.
- Pickering, M. J. & Garrod, S. (2004). Toward a mechanistic psychology of dialogue. *Behavioral and Brain Sciences*, 27(2), 169–190. <http://dx.doi.org/10.1017/S0140525X04000056>
- Pickering, M. J. & Garrod, S. (2006). Alignment as the basis for successful communication. *Research on Language and Computation*, 4 (2–3), 203–228. <http://dx.doi.org/10.1007/s11168-006-9004-0>
- Raveh, E., Siegert, I., Steiner, I., Gessinger, I., & Möbius B. (2019). Three's a crowd? Effects of a second human on vocal accommodation with a voice assistant. *Interspeech 2019*, Graz, 4005–4009. <http://dx.doi.org/10.21437/Interspeech.2019-1825>
- Reitter, D., Moore, J. D., & Keller, F. (2006). Priming of syntactic rules in task-oriented dialogue and spontaneous conversation. In R. Sun (Ed.), *Proceedings of the 28th Annual Conference of the Cognitive Science Society* (pp. 685–690). Mahwah: Lawrence Erlbaum Associates, Inc.
- Ross, J. P., Lilley K. D., Clopper, C. G., Pardo, J. S., & Levi, S. V. (2021). Effects of dialect-specific features and familiarity on cross-dialect phonetic convergence. *Journal of Phonetics*, 86, 101041. <http://dx.doi.org/10.1016/j.wocn.2021.101041>
- Ruch, H. (2015). Vowel convergence and divergence between two Swiss German dialects. *18th International Congress of Phonetic Sciences*, Glasgow, UK.
- Ruch, H. (2018). The role of acoustic distance and sociolinguistic knowledge in dialect identification. *Frontiers in Psychology*, 9, 818. <http://dx.doi.org/10.3389/fpsyg.2018.00818>
- Ruch, H., Zürcher Y., & Burkart J. (2017). The function and mechanism of vocal accommodation in humans and other primates. *Biological Reviews*. <http://onlinelibrary.wiley.com/doi/10.1111/brv.12382/full>.



- Sancier, M. L. & Fowler, C. A. (1997). Gestural drift in a bilingual speaker of Brazilian Portuguese and English. *Journal of Phonetics*, 25(4), 421–436. <http://dx.doi.org/10.1006/jpho.1997.0051>
- Sanker, C. (2015). Comparison of phonetic convergence in multiple measures. *Cornell Working Papers in Phonetics and Phonology 2015*, 60–75.
- Schweitzer, A. & Lewandowski, N. (2014). Social factors in convergence of F1 and F2 in spontaneous speech. *International Seminar on Speech Production*, Cologne. <https://www.ims.uni-stuttgart.de/documents/team/schweitz/docs/SchweitzerLewandowski2014.pdf>
- Shockley, K., Sabadini, L., & Fowler, C. A. (2004). Imitation in shadowing words. *Perception & Psychophysics*, 66(3), 422–429. <http://dx.doi.org/10.3758/BF03194890>
- Soderstrom M. (2007). Beyond babytalk: Re-evaluating the nature and content of speech input to preverbal infants. *Developmental Review*, 27(4), 501–532. <http://dx.doi.org/10.1016/j.dr.2007.06.002>
- Soliz, J. & Giles, H. (2016). Relational and identity processes in communication: A contextual and meta-analytical review of Communication Accommodation Theory. *Annals of the International Communication Association*, 38(1), 107–144. <http://dx.doi.org/10.1080/23808985.2014.11679160>
- Van Engen, K. J., Baese-Berk, M., Baker, R. E., Choi, A., Kim, M., & Bradlow, A. R. (2010). The Wildcat Corpus of native-and foreign-accented English: Communicative efficiency across conversational dyads with varying language alignment profiles. *Language and Speech*, 53(4), 510–540. <http://dx.doi.org/10.1177/0023830910372495>
- Walker, A. & Campbell-Kibler, K. (2015). Repeat what after whom? Exploring variable selectivity in a cross-dialectal shadowing task. *Frontiers in Psychology*, 6(546). <http://dx.doi.org/10.3389/fpsyg.2015.00546>
- Walters, S. A., Babel, M. E., & McGuire, G. (2013). The role of voice similarity in accommodation. *Proceedings of Meetings on Acoustics*, 19(1), 060047.58. <http://dx.doi.org/10.1121/1.4800716>
- Ward, A. & Litman, D. (2007). Automatically measuring lexical and acoustic/prosodic convergence in tutorial dialogue corpora. In *SLaTE Speech and Language Technology in Education 2007*.
- White, L. & Mattys, S. L. (2007). Calibrating rhythm: First language and second language studies. *Journal of Phonetics*, 35(4), 501–522. <http://dx.doi.org/10.1016/j.wocn.2007.02.003>
- Zellou, G., Scarborough, R., & Nielsen, K. (2016). Phonetic imitation of coarticulatory vowel nasalization. *The Journal of the Acoustical Society of America*, 140(5), 3560–3575. <http://dx.doi.org/10.1121/1.4966232>