SPEAKER-IDIOSYNCRASY IN PAUSING BEHAVIOR: EVIDENCE FROM A CROSS-LINGUISTIC STUDY

Marie-José Kolly[a,b], Adrian Leemann[c], Philippe Boula de Mareüil[a], Volker Dellwo[b]

[a]LIMSI-CNRS Orsay, [b]Phonetics Laboratory, Department of Comparative Linguistics, University of Zurich, [c]Phonetics Laboratory, Department of Theoretical and Applied Linguistics, University of Cambridge
{marie-jose.kolly|volker.dellwo}@uzh.ch, al764@cam.ac.uk, philippe.boula.de.mareuil@limsi.fr

ABSTRACT

Phoneticians study acoustic speech signals. But what about the aspects of speech where the signal is silent? The present study investigated speakers' pausing behavior in their native and non-native speech. Pausing measures were applied in order to study between-speaker and within-speaker variability, where within-speaker variability was introduced by recording speakers in their native Zurich German, and in their second languages English and French. Results showed that pausing measures in the form of pause numbers and pause durations are speaker-specific. Furthermore, this speaker-specificity became evident across different languages. Results are discussed in the context of forensic voice comparison.

Keywords: pausing, temporal features, speaker-idiosyncrasy, second language, forensic phonetics

1. INTRODUCTION

Speakers, native and non-native, produce silent pauses when they speak or read aloud. Such pauses can occur in places where a pause is allowed by the syntactic makeup of the sentence – or elsewhere, where they may be perceived as "disfluencies" [18].

In the past, non-native speakers' pausing behavior was often investigated as a correlate of perceived fluency [4, 5] or as an indicator of second language proficiency [24, 25]. [6] note that pausing behavior also has to do with personality or style.

The experiment reported in the present paper was designed to explore speaker-specific pausing: two non-native speakers with the same language background and similar second language proficiency might have different habits or preferences regarding the frequency and duration of silent pauses in their speech – be it L1 or L2 speech. If this is the case, then pausing behavior may be an interesting measure for the domain of forensic phonetics. In typical cases of forensic voice comparison, trace material from a crime – e.g. recordings of a perpetrator of a bomb threat – is compared to acoustic comparison material – e.g. recordings of a suspect during a police interview – and used in forensic investigations. Acoustic measures that vary between speakers but are invariant within speakers, i.e. speaker-specific measures, are thus desirable for applications in forensic speaker comparison [22].

Speaker-specific behavior exists in different types of acoustic features. Research has revealed between-speaker variability in the frequency domain – in formant frequencies [19, 20] and fundamental frequency [13, 16, 21] –, and in the intensity domain [1]. Only recently has research shown speaker-idiosyncratic patterns in the time domain: [7, 8, 15–17] found suprasegmental temporal features to be speaker-specific and robust to within-speaker variability. Within-speaker variability introduced in forensic phonetic studies typically includes speaking style variability (read vs. spontaneous speech [8, 14–16]), channel variability (hifi vs. telephone speech, [14, 15]), and voice disguised speech [13].

Do speakers differ in their pausing behavior? And does pausing behavior remain speaker-specific if speakers talk in different languages? We introduced between-speaker variability by studying 16 speakers, and included within-speaker variability by having the same speakers produce native Zurich German speech, and non-native English and French speech.

2. METHODS

2.1. Speakers

16 speakers of Zurich German (eight male / eight female) were recorded at the Phonetics Laboratory, University of Zurich, to create the TEVOID corpus [7, 8, 15–17]. Speakers' age ranged between 20 and 33 years (M=25.4; SD=3.7). All speakers were University of Zurich students who spoke the dialect of the city of Zurich. They thus showed little to no regional accent variability, as attested by informal listening tests. All speakers had learned French and English as a second language at school. Usually, French classes started at age 8 and English classes at age 13. The speaker group was thus relatively homogeneous in terms of native dialect, age, education and second languages spoken. Recordings were made in a sound-treated booth using an omnidirectional Earthworks QTC40 high definition

condenser microphone (sampling rate of 44.1kHz; 16 bit quantization). Speakers were paid 30 Swiss Francs per hour for their participation.

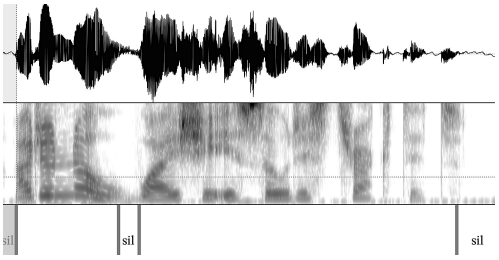
2.2. Material

Each speaker read 16 Zurich German sentences, 16 English sentences and 16 French sentences taken from the TEVOID corpus. English and French sentences were literal translations of the Zurich German sentences (yet idiomatic in English and French) and were thus roughly similar in length: sentences typically contained 15–20 syllables. These 768 sentences (16 speakers × 16 sentences × 3 languages) constituted the corpus used in the present study. Prior to the recording, speakers had prepared reading the sentences at home, to ensure fluent reading of the material. If hesitations in the form of filled pauses occurred in a sentence, speakers repeated the sentence spontaneously or, if not, they were asked to do so. Sentences which contained hesitations in the form of silent pauses were not repeated, however.

2.3. Data editing

To prepare the data for the application of pausing measures (cf. 2.4), trained phoneticians (first and second author) labeled each sentence for silent pauses using Praat software [3]. Speakers may pause to reflect syntactic constituents in spoken language, e.g. between a main and a subordinate clause, to mark conversational structure, e.g. emphasize a subsequent stretch of speech, or for stylistic reasons, e.g. to reflect idiolectal aspects of speech. In addition, speakers may pause for cognitive reasons, e.g. hesitating as a means to prepare for what to say next. All these types of silent pauses were labeled in our corpus, which means that no duration threshold was applied for the labeling of silent parts. Pauses were labeled perceptually – every silent part which was perceived as a pause was labeled as such (indicated by the interval label *sil* in Figure 1). Every sentence in the corpus is preceded and followed by a (labeled) pause, cf. Figure 1.

Figure 1: Praat TextGrid with hand-labeled pauses in the English sentence *I don't know* [pause] *why she is so distracted*.



2.4. Pausing measures applied

We applied two measures that describe speakers' pausing behavior and are widely used in second language research [4–6, 9, 10, 14, 18, 24–27]:

1. The number of pauses in a sentence: *pauseNbr*.
2. The sum of the durations (in seconds) of all pauses in a sentence: *pauseDur*.

The silences that precede and follow each sentence were not taken into account for the calculation of *pauseNbr* and *pauseDur*.

2.5. Speech tempo effects

Findings of [9, 26, 27] suggest that, for some speakers, pausing behavior covaries with articulation rate. We therefore checked whether *pauseNbr* or *pauseDur* may be influenced by articulation rate. As a measure of articulation rate, we calculated *ratePeak*: the number of automatically detected peaks in the amplitude envelope – which roughly corresponds to the number of syllables – per second, excluding pauses [7]. Neither *pauseNbr* (r=0.16) nor *pauseDur* (r=0.10) were correlated with *ratePeak*.

2.6. Statistical analyses

Data were analyzed using linear mixed effect models (LMEs), with R software [23] and the R package *lme4* [2]. *Language* was included as a fixed effect, *speaker* and *sentence* as random effects. We included a random slope of *speaker* on *language* to test for interactions between the two factors. Effects were tested by model comparison between a full model in which the factor in question was present and a reduced model in which the factor was excluded. We applied standard likelihood ratio tests to compare the two models. We report AIC (Akaike Information Criterion) values for the relative goodness of fit [12]. We assumed an α level of 0.01.

3. RESULTS

3.1. Number of pauses: *pauseNbr*

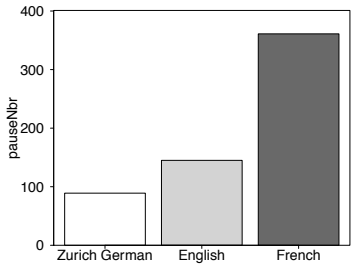
Table 1 summarizes the results obtained for *pauseNbr*. The AIC values are equal for each test because they are based on the full model, which, for every factor, provided an improved goodness of fit.

Table 1: Summary of the LMEs for *pauseNbr*

Factor	Result
language	p<0.0001; AIC=1740
speaker	p<0.0001; AIC=1740
*language*speaker*	p<0.0001; AIC=1740
sentence	p<0.0001; AIC=1740

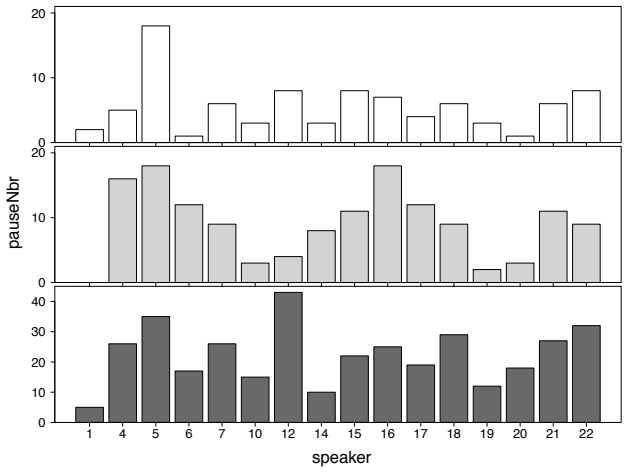
language was found to be highly significant, cf. Figure 2: *pauseNbr* was lowest in Zurich German (M=0.35, SD=0.63), followed by English (M=0.57, SD=0.69) and French (M=1.41, SD=1.22).

Figure 2: Barplots of *pauseNbr* per *language* for Zurich German, English, and French.



We also found a highly significant effect of *speaker*, cf. Figure 3. Since there was a significant interaction between *language* and *speaker*, we calculated simple effects for the factor *speaker* on the Zurich German, English and French data separately (Bonferroni corrected α: 0.01/3=0.003). *speaker* was significant in the Zurich German ($p<0.0001$; AIC=431), English ($p<0.0001$; AIC=508) as well as in the French ($p<0.0001$; AIC=730) data. We also calculated simple effects for the factor *language* for each speaker separately (Bonferroni corrected α: 0.01/16=0.0006). *language* was only significant in 7 out of 16 speakers. Furthermore, *pauseNbr* was affected by the highly significant factor *sentence*.

Figure 3: Barplots of *pauseNbr* by *speaker* for Zurich German (top, white), English (center, light gray), and French (bottom, dark gray). NB: y-axes show different maxima.



3.2. Pause durations: *pauseDur*

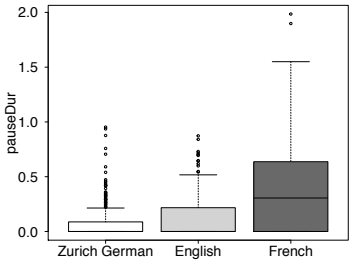
Table 2 summarizes the results obtained for *pauseDur*.

Table 2: Summary of the LMEs for *pauseDur*.

Factor	Result
language	$p<0.0001$; AIC=-139
speaker	$p<0.0001$; AIC=-139
*language*speaker*	$p<0.0001$; AIC=-139
sentence	$p<0.0001$; AIC=-139

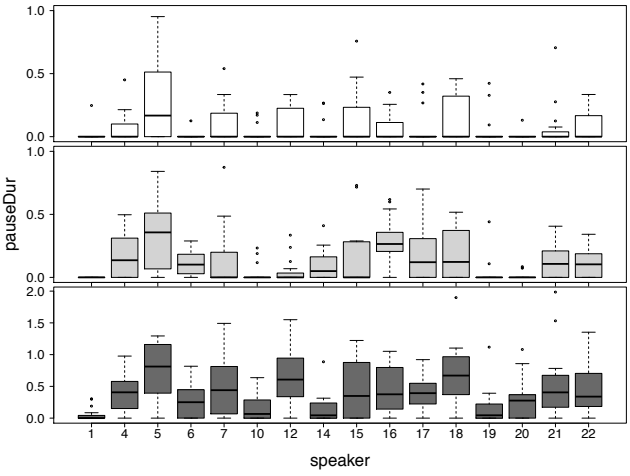
language was highly significant, cf. Figure 4: *pauseDur* was lowest in Zurich German (M=0.08, SD=0.17), followed by English (M=0.13, SD=0.18) and French (M=0.40, SD=0.41). There was a highly significant effect of *speaker*, cf. Figure 5. Since the interaction of *language* and *speaker* was significant, we calculated simple effects for *speaker* as described in 3.1. *speaker* was significant in Zurich German ($p<0.0001$; AIC=-244), English ($p<0.0001$; AIC=-174) as well as French ($p<0.0001$; AIC=-142).

Figure 4: Boxplots of *pauseDur* by *language* for Zurich German, English, and French.



We also calculated simple effects for *language* (cf. 3.1). Again, *language* was only significant in 7 out of 16 speakers. Furthermore, *pauseDur* was affected by the highly significant factor *sentence*.

Figure 5: Boxplots of *pauseDur* by *speaker* for Zurich German (top, white), English (center, light gray), and French (bottom, dark gray). NB: y-axes show different maxima.



4. DISCUSSION

In terms of *language* effects, we found that speakers produced the fewest and the shortest pauses in their native Zurich German speech, and the most and the longest pauses in their French speech. Speakers' pausing behavior in English was located in between French and German. This may be explained by the cognitive task at hand: speaking a second language is cognitively more demanding than speaking a first language – in which speakers are more proficient: [10] has shown that cognitively more demanding tasks lead to longer pauses in speech. This is corroborated by [24], who shows that second language proficiency affects the number and duration of pauses. The difference between the two second languages French and English in our data is most likely explained by the fact that Zurich German speakers are more proficient in English than in French. Even though, at school, they learned French before English and even though they live in a country where French is an official language, Swiss German university students most probably hear and produce English more often than French.

We found an effect of *sentence* for the number and the duration of pauses. Looking at the data more closely, results showed that certain sentences – such as English (E1): *One could either help serving in the house, or go outside* and French (F1): *Soit on aidait à servir là-bas, dans la maison, soit on allait dehors* – show many and long pauses. In (E2): *I am really interested in everything* and (F2): *Je suis vraiment intéressée à tout*, there were fewer and shorter pauses. (E1) and (F1) are some of the longest sentences of the corpus, and thus are more likely to show pauses because of that. Furthermore, their syntactic construction provides potential slots for pauses that co-occur with punctuation such as commas. (E2) and (F2), on the other hand, are very short and made up of one main clause only. Fewer pauses are thus expected in these sentences.

A higher number of pauses is expected to lead to more occurrences of phrase-final lengthening and thus to a lower articulation rate. This was not the case in our corpus: number and duration of pauses were not related to our measure of articulation rate. This finding may be due to the – possibly too coarse – peak detection method applied, which leaves room for further investigations in the future.

In terms of *speaker* effects, our data revealed significant between-speaker differences, in the number as well as the duration of pauses. At the same time, measures varied little within speakers: only for 7 out of 16 speakers did we observe a simple effect of *language*. Speaker 1, for example, made few pauses in Zurich German, French as well as in English speech. Speaker 5, on the other hand, showed high values for the number of pauses in all three languages. The same holds for pause durations: speaker 1 produced short pauses in all three languages, whereas speaker 5 produced long pauses.

As for the implications for the domain of forensic phonetics, both pausing measures showed significant between-speaker variability on the one hand and little within-speaker variability on the other. When testing simple effects of *language*, 7 out of 16 speakers did not differ in their pausing behavior – regardless of whether they spoke Zurich German, English or French. Furthermore, Figures 3 and 5 show that, even if there was an effect of *language* for a particular speaker, the direction of the effect was most often constant: speakers produced most and the longest pauses in French and least and the shortest pauses in Zurich German. This is surprising, since [14] found low speaker-specific values for speakers' pausing behavior, whereas within-speaker variability – introduced by having speakers read and speak spontaneously – was relatively high.

The International Association for Forensic Phonetics and Acoustics (IAFPA, [11]) advises members to "exercise particular caution" when carrying out analyses on non-native speech. More extensive research about L2 speech may complement existing parameters that are used in forensic casework. More importantly, incriminating speech samples are frequently recorded over a telephone, which degrades the spectral characteristics of the acoustic signal but does not affect temporal characteristics such as pausing [14].

5. CONCLUSION AND FUTURE WORK

The present study set out to investigate whether pausing behavior is speaker-specific, and the degree to which this is true across different languages. Results showed high between- and low within-speaker variability in the number and duration of pauses in each sentence. This suggests that temporal measures such as speakers' pausing behavior may be useful for the domain of forensic voice comparison. Further steps in this research will include an increase in size of the database and the application of a wider variety of temporal measures, cf. [7, 8, 15–17].

6. ACKNOWLEDGEMENTS

This research was supported by the Swiss National Science Foundation (SNSF; grant numbers 135287 and 155024). We would like to thank Bob Ladd for valuable feedback on a first version of this manuscript and Stephan Schmid for his expert advice on foreign-accented speech.

7. REFERENCES

[1] Amino, K., Arai, T. 2009. Speaker-dependent characteristics of the nasals. *Forensic Science International* 185, 21–28.

[2] Bates, D.M., Maechler, M. 2009. lme4: Linear mixed-effects models using S4 classes. R package version 1.1-7.

[3] Boersma, P., Weenink, D. 2012. *Praat: Doing phonetics by computer*. http://www.praat.org/.

[4] Bosker, H. R., Quené, H., Sanders, T., Jong, N. H. 2014. The perception of fluency in native and nonnative speech. *Language Learning* 64, 579–614.

[5] Cucchiarini, C., Strik, H., Boves, L. 2002. Quantitative assessment of second language learners' fluency: Comparisons between read and spontaneous speech. *Journal of the Acoustical Society of America* 111, 2862–2873.

[6] de Jong, N. H., Groenhout, R., Schoonen, R., Hulstijn, J. H. 2013. Second language fluency: Speaking style or proficiency? Correcting measures of second language fluency for first language behaviour. *Applied Psycholinguistics* 34, 1–21.

[7] Dellwo, V., Leemann, A., Kolly, M.-J. 2012. Speaker idiosyncratic rhythmic features in the speech signal. *Proceedings of Interspeech 2012*, Portland, USA.

[8] Dellwo, V., Leemann, A., Kolly, M.-J. (2015). Rhythmic variability between speakers: Articulatory, prosodic and lexical factors. To appear in: *Journal of the Acoustical Society of America* 137, 1513–1528.

[9] Fougeron, C., Jun, S.-A. 1998. Rate effects on French intonation: Prosodic organization and phonetic realization. *Journal of Phonetics* 26, 45–69.

[10] Grosjean, F. 1980. Temporal variables within and between languages. In: Dechert, H. W., Raupach, M. (eds), *Towards a Cross-Linguistic Assessment of Speech Production*. Frankfurt: Lang, 39–53.

[11] IAFPA = International Association for Forensic Phonetics and Acoustics. http://www.iafpa.net.

[12] Kliegl, R., Wei, P., Dambacher, M., Yan, M., Zhou, X. 2011. Experimental effects and individual differences in linear mixed models: Estimating the relationship between spatial, object, and attraction effects in visual attention. *Frontiers in Psychology* 1, 1–12.

[13] Künzel, H. J. 2000. Effects of voice disguise on speaking fundamental frequency. *Forensic Linguistics* 7, 149–179.

[14] Künzel, H. J. 2013. Some general phonetic and forensic aspects of speaking tempo. *International Journal of Speech Language and the Law* 4, 48–83.

[15] Leemann, A., Kolly, M.-J., Dellwo, V. 2014. Speaker-individuality in the time domain: Implications for forensic voice comparison. *Forensic Science International* 238, 59–67.

[16] Leemann, A., Mixdorff, H., O'Reilly, M., Kolly, M.-J., Dellwo, V. (2015). Speaker-individuality in Fujisaki model f0 features: Implications for forensic voice comparison. *International Journal of Speech, Language and the Law* 21, 343–370.

[17] Leemann, A., Kolly, M.-J., Dellwo, V. (in review). Speaker-invariant suprasegmental temporal features in normal and disguised speech.

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

[19] McDougall, K. 2004. Speaker-specific formant dynamics: An experiment on Australian English /aI/. *International Journal of Speech, Language and the Law* 11, 103–130.

[20] Morrison, G. 2009. Likelihood-ration based forensic speaker comparison using representations of vowel formant trajectories. *Journal of the Acoustical Society of America* 125, 2387–2397.

[21] Nolan, F. 2002. Intonation in speaker identification: An experiment on pitch alignment features. *Forensic Lingistics* 9, 1–21.

[22] Nolan, F. 2009. *The phonetic bases of forensic speaker identification*. 2nd ed. Cambridge: Cambridge University Press.

[23] R Core Team 2013. R. A language and environment for statistical computing. Version 3.0.1. Vienna. http://www.R-project.org.

[24] Riazantseva, A. 2001. Second language proficiency and pausing. A study of Russian speakers of English. *Studies in Second Language Acquisition* 23, 497–526.

[25] Trofimovich, P., Baker, W. 2006. Learning second language suprasegmentals: Effect of L2 experience on prosody and fluency characteristics of L2 speech. *Studies in Second Language Acquisition* 28, 1–30.

[26] Trouvain, J. 2003. *Tempo variation in speech production. Implications for speech synthesis*. PhD Thesis, University of Saarbrücken.

[27] Trouvain, J., Grice, M. 1999. The effect of tempo on prosodic structure. *Proceedings of the 14th ICPhS*, San Francisco, 1067–1070.