Speed, breakdown, and repair: An investigation of fluency in long-term L2 speakers of English

Author Note

Cornelia Lahmann, English Department, University of Groningen

Rasmus Steinkrauss, English Department, University of Groningen

Monika Schmid, Department of Language and Linguistics, University of Essex/English Department, University of Groningen

This research was financed by the Dutch national research foundation (NWO; Nederlandse Organisatie voor Wetenschappelijk Onderzoek), Grant No. 360-70-420. We also thank the following archives for granting us the permission to use their oral history testimony data: USC Shoah Foundation Institute (Los Angeles), *Werkstatt der Erinnerungen* (Hamburg), *Alte Synagoge* (Essen), Fortunoff Video Archive for Holocaust Testimonies (Yale University Library), Tauber Holocaust Library and Education Program (San Francisco), United States Holocaust Memorial Museum (Washington, D.C.), Association of Jewish Refugees (London).

Correspondence concerning this article should be addressed to Cornelia Lahmann, University of Groningen, English Department, Faculty of Arts, P.O. Box 716, NL 9700AS Groningen

Email: cornelia.lahmann@gmx.de

Abstract

Aims and Objectives/Purpose/Research Questions

The present study investigated which factors would best predict L2 fluency in a group of long-term L2 speakers of different English varieties with German as their L1.

Design/Methodology/Approach

L2 fluency was conceptualized in terms of utterance fluency for which speed, breakdown, and repair fluency were distinguished.

Data and Analysis

Multiple measures of utterance fluency were applied to four-minute speech fragments originating from 102 spontaneous oral interviews. Interviewees' ages of onset ranged from 7 to 17 whereas their ages at interview ranged from 57 to 87. Multifactorial analyses yielded significant effects of age at interview.

Findings/Conclusions

Whereas the mean number of silent pauses and repairs increases, syllable duration decreases. This leaves room for interpretation as to why we find an aging effect. Overall, the evidence suggests that the usual, SLA-specific factors such as age of onset or length of residence are no longer at play to predict L2 fluency.

Originality, and Significance/Implications

To this point L2 fluency in very advanced, highly proficient L2 speakers has received little attention. The results point to the need for more research into highly proficient L2 users.

Fluency is a concept that to some extent eludes a straightforward definition and measurement, and different approaches abound (Chambers, 1997; Fillmore, 1979; Lennon, 1990, 2000; Pawley & Syder, 1983; Rehbein, 1987). It is generally considered a characteristic of first language (L1) speech production by healthy adult native speakers, and it seems to be an indicator of second language (L2) proficiency. As one of the interviewees in our study pointed out with respect to acquiring L2 proficiency: “Within a few months I was fluent in English”. This broad conception of fluency refers to global oral proficiency (Lennon, 1990). With regard to second language proficiency it extends to the L2 speaker’s high command of the foreign or second language (Ginther, Dimova, & Yang, 2010). However, it is fluency in the narrow sense in terms of speech production which will be the focus of the present investigation (see e.g., De Jong, Steinel, Florijn, Schoonen, & Hulstijn, 2013; Kormos & Dénes, 2004; Segalowitz, 2010; Skehan, 2003). In accordance with previous definitions postulated by Rehbein (1987) and Schmidt (1992), fluency in the narrow sense is defined as “the ability to produce the L2 with nativelike rapidity, pausing, hesitation, or reformulation” (Housen, Kuiken, & Vedder, 2012, p. 2). Hence, disfluencies resulting from pausing, hesitation, or reformulation are a natural part of speech, for L1 and L2 speakers alike.

Where L2 speakers deviate from the native norm (whether for fluency or other types of L2 skills), it is often attributed to lower levels of proficiency, in particular in late-onset L2 learners. In addition to age of onset, alternative variables such as language use and socio-psychological variables like motivation toward the L2 have been suggested to explain different degrees of L2 achievement more generally. While these factors have been widely explored for several domains of L2 knowledge, such as grammar (Birdsong & Molis, 2001; Johnson & Newport, 1989, 1991) and phonology (e.g., Oyama, 1976; Piske et al., 2001), we still know fairly little about the fluency of highly advanced L2 users (in comparison with L1 users' fluency) (Forsberg Lundell et al., 2013; Götz, 2013; Kormos & Dénes, 2004), in particular for those who spent the majority of their lives in the L2 community. This paper therefore addresses which factors affected the L2 fluency/ disfluency of long-term German-Jewish immigrants to various English-speaking countries.

Theoretical background

(Dis)Fluency in L1 and L2 speech. Since L2 fluency is linked to L1 fluency in the narrow definition of the term, let us first take a look at L1 fluency. In his blueprint for the healthy adult native speaker, Levelt (1989, p. 9) illustrates the various processing components and knowledge stores involved in the intentional production of speech. The *Conceptualizer* constitutes the processing system in which the intention to talk is conceived and the relevant information for expressing oneself is selected and ordered. The *Formulator* receives the message that has previously been conceptualized and outputs a phonetic or articulatory plan for which assessing the lexicon as a crucial knowledge store is necessary. The output of the formulator becomes the input to the *Articulator*, which executes the articulatory plan. Since a speaker is its own listener, he can attend not only to other people's speech but also his own internal and overt speech. Self-monitoring is therefore a crucial characteristic of Levelt's blueprint.

If the incremental processes moving from conceptualizer through formulator and articulator are not interrupted and if no 'flaws' are detected during the monitoring, then the speaker produces fluent speech. If the reverse is the case, we will find disfluencies, i.e. pauses and self-repairs. Disfluencies are a 'normal' part of speech produced by any native speaker. A certain degree of disfluency does usually not interrupt the speech stream and is also not perceived as deviant by listeners.

Significant increases in disfluency are often observed in groups whose access to the speaking production process as described by Levelt is less automated than for healthy adult native speakers. One such group may be L2 speakers, which produce significantly more disfluencies as compared to native speakers (Cucchiarini, Strik, & Boves, 2000). Different proposals have been put forth to explain L2 speakers' disfluencies. First of all, L2 speakers face the challenge of having to access a second language in addition to their first. A bilingual adaptation of Levelt's model has been postulated by de Bot (1992) to illustrate this special, bilingual 'condition'. It shows that where language-specific processing is involved, i.e. during the second phase in the conceptualizer and during lexical access at the level of the formulator, the existence of a second language can lead to breakdown and possibly the need to repair. De Bot attributes the slowing down in bilinguals to the simultaneous activation of both languages, where a decrease in speech rate is to be expected for the weaker language. Complementary, Lennon (2000, p. 32) states four possible causes for L2 speakers' increased levels of disfluency as found e.g. by Elsendoorn (1984), Flege (1979), Munro & Derwing (1995), and Towell, Hawkins, and Bazergui, (1996): (1) Articulation in the L2 is less well automatized; (2) there is competition from the L1 with regard to the formulation process, (3) less proficient speakers have to rely on controlled instead of automated processes for accessing and formulating, (4) and there may be deficits in linguistic storage. Altogether, there may be a number of possible underlying reasons for L2 speakers' disfluencies, ranging from different processing mechanisms, and competition between L1 and L2, to proficiency. In turn these causes may also be interrelated, at lower levels of proficiency more controlled processing is to be expected as well as a higher degree of competition.

In order to make empirical claims and to generalize about possible causes for L2 disfluency, agreement needs to be reached regarding the operationalization and assessment of fluency to ensure comparable investigations e.g., on the effects of age of onset or continued L1 exposure on fluency.

Operationalization and the assessment of a multidimensional L2 fluency. Fluency is considered to be a multi-level, multidimensional construct. At the level of utterance fluency (Segalowitz, 2010), which comprises the specific characteristics that a speech sample can possess, a three-way distinction between speed, breakdown, and repair fluency was proposed by Skehan (2003). It suggests utterance fluency to be a multidimensional construct for which several measures have been suggested.

A number of studies investigated the relationship between fluency and proficiency with the aim to find out which measures of (dis)fluency best predicted oral proficiency. Based on a review of 12 studies[1] we found that several measures of breakdown, speed, and repair according to Skehan's distinction correlated well with either subjective ratings of fluency in particular or ratings of oral proficiency in general. The following measures of breakdown correlated particularly well with proficiency measures: number of silent pauses, length of silent pauses, number and length of mid-clause pauses (Bosker et al., 2013; Cucchiarini et al., 2000; Cucchiarini, Strik, & Boves, 2002; Ginther, Dimova, & Yang, 2010; Iwashita, 2010; Iwashita, Brown, McNamara, & O'Hagan, 2008). For speed fluency measures related to syllable counts, e.g., the length of syllables, the mean number of syllables and the mean number of pruned syllables correlated highly with proficiency measures (Bosker et al., 2013; Cucchiarini et al.,

[1] (Bosker, Pinget, Quené, Sanders, & De Jong, 2013; Cucchiarini, Strik, & Boves, 2000; Nivja H. De Jong, Steinel, Florijn, Schoonen, & Hulstijn, 2013; Derwing, Rossiter, Munro, & Thomson, 2004; Gelderen, 1994; Ginther, Dimova, & Yang, 2010; Iwashita, 2010; Iwashita, Brown, McNamara, & O'Hagan, 2008b; Kormos & Dénes, 2004; Lennon, 1990; Mizera, 2006; Rossiter, 2009)

2000, 2002; Derwing et al., 2004; Ginther et al., 2010; Iwashita, 2010; Iwashita, Brown, McNamara, & O'Hagan, 2008; Kormos & Dénes, 2004; Mizera, 2006; Rossiter, 2009). As for repair fluency, only a few studies showed that repairs such as repetitions, replacements, false starts, reformulations, and so-called nonfunctional pauses are related to oral proficiency (Gelderen, 1994; Iwashita, 2010; Rossiter, 2009). In sum, these are the type of fluency measures most frequently used in L2 acquisition research (in particular for assessing L2 development) and which have been shown to reliably predict L2 proficiency.

On the basis of this type of operationalization of fluency, a number of experimental SLA studies have explored the relationship between L1 and L2 fluency and possible predictors of L2 disfluency. Such studies found e.g., that a speaker's L2 fluency is related to his L1 fluency (De Jong, Groenhout, Schoonen, & Hulstijn, 2013; Derwing, Munro, Thomson, & Rossiter, 2009). A few studies have compared fluency measurements of non-native speakers with those of native speakers mostly in terms of the duration of utterances spoken in an L2. Findings show that late-onset L2 learners produce longer utterances in terms of duration (e.g., Elsendoorn, 1984; Flege, 1979; Munro & Derwing, 1995); and that L2 proficiency affects L2 fluency, e.g. pausing patterns (Riazantseva, 2001). In sum, studies found that L2 speakers pause longer and more frequently, but also in different locations than is the case for natives. However, such disfluencies appear to be modulated by different factors, such as L1 fluency, L2 proficiency or age of onset.

The age of onset factor has received particularly great attention in L2 ultimate attainment research in the domains of L2 grammar, lexicon and phonology (Abrahamsson & Hyltenstam, 2009; Granena & Long, 2013b; Hellman, 2011). Such studies usually look at situations of naturalistic L2 acquisition. To our knowledge there are only a few studies which have looked at age of onset effects on temporal aspects of fluency in a natural L2 setting with immigrant populations albeit not necessarily at the latest stage(s) of L2 development with a maximum length of residence (Derwing & Munro, 2013; Guion, Flege, Liu, & Yeni-Komshian, 2000; Mackay & Flege, 2004; Trofimovich & Baker, 2006). In these studies participants with various L1 backgrounds including Italian, Korean, Slavic languages, and Mandarin learned English as a second language either in the US or Canada. Guion et al. (2000) studied learners with AOs ranging from 3 to 22 and a mean LoR of 32. Mackay & Flege (2004) compared a group of early bilinguals with AOs from 2 to 13 years and late bilinguals with AOs from 15 to 28 years. Both studies found that sentence duration increased the later the AO, whereas younger learners' mean sentence duration was comparable to those of native speakers. Trofimovich & Baker (2006) report on a study of adult L2 learning in which they looked at three learner groups ranging from inexperienced to experienced. The group considered to be inexperienced had a mean age of onset of 29 and a mean length of residence of 3 months (in the U.S.). The second group had a mean age of onset of 24 with a mean length of residence of 3 years. The experienced group had a mean age of onset of 21 with a mean length of residence of 10 years. They found that age of onset was the primary predictor of speech rate, as well as pause duration and frequency, also when LoR was partialled out. Altogether, even the most experienced L2 learners were different from native controls. The study suggests an influence of age of onset in adulthood and, thus, beyond previously hypothesized critical periods. Similar evidence comes from Derwing and Munro (2013) who looked at adult immigrant learners with AOs ranging from 19 to 49 and LoRs of 7 years (Derwing, Munro, & Thomson, 2008) with either Mandarin or a Slavic language as their L1. Derwing and Munro (2013) concluded that comprehensibility, fluency, and accent improvement were affected by an interplay of L1, age of onset, the depth and breadth of learners' conversations in English, and their willingness to communicate. In sum, the studies referred to above looked at groups varying in age of onset, age at testing, and length of residence. They generally suggest that learners with earlier ages of onset are more likely to reach nativelike fluency. However the number of studies looking at situations of long-term L2 exposure is limited (Guion et al., 2000; Mackay & Flege, 2004).

Research questions

Previous studies suggest that age of onset may contribute to significant increases in L2 disfluencies, but other factors such as L1 fluency might also be related. However, the majority of these studies looked at L2 speakers with limited LoRs (from a few months up to 10 years). The question remains whether fluency is affected by age of onset in situations of long-term residency where speakers, who were exposed to the L2 sufficiently early in life, spent more than half of their lives in the L2 community. In our study we set out to probe this question by looking at a group of long-term L2 speakers with ages of onset from 7 to 17. The following research question will be addressed:

- RQ1: How does age of onset affect L2 fluency in spontaneous oral production of long-term L2 speakers?
- RQ2: Are there any other variables that significantly predict their L2 fluency?

Methods

Interviewees. The study is comprised of 102 oral history interviews. The interviewees' mean age of onset (AO, which equals their age at emigration) was 12.2 years with a mean length of residence (LoR) of 61.3 years. Their mean age at the time of the interview (AaI) was 73.6. The majority of participants were female (n = 60). They were assessed with respect to their level of continued exposure to their L1 German (L1 Exp) after emigration. Three independent raters were asked to rate the interviewees' L1 Exp on a scale from 1 (low) to 7 (high). Their final ratings were based on the occurrence of statements about (a) avoidance of speaking German, (b) manner of emigration (adoption into foster family or not), (c) contact with family members, (d) origin of marital partner (native German, native English, or other nationality), (e) continued use of German (during studies, work, or extracurricular engagements), and (f) integration into English-speaking community (through studies, work, and/or extracurricular engagements). Interrater agreement for all pairs was $r \geq .7$. We considered the median value as the final L1 Exp score to avoid outliers. Interviewees' average L1 Exp was 4.34. In addition to L1 Exp we also include a categorical variable for use of German at work. This variable is more concrete than L1 Exp. In addition, it is likely to complement the L1 exposure variable in that the use of German at work is expected to be of higher quality as compared to using it in an everyday life family context. Opportunities for switching may also be reduced in a working environment, allowing for monolingual (L1 German) mode (e.g., Schmid & Dusseldorp, 2010). Based on the interviews it was established that 14 interviewees used German at work and 66 did not. For 22 interviewees information on the use of German at work was not available. In terms of level of education (Edu), 46 participants graduated from university (high Edu) and 34 had obtained a high school diploma and possibly some additional vocational training (medium Edu). 11 of the participants had not completed their high school education (low Edu). For the remaining 11 participants we did not have any information on their Edu. See Table 1 for a summary of the independent background variables.

[Table 1]

Oral histories. The data were oral history testimonies, i.e. personal narratives in which our interviewees narrated their lives both before and after emigration. The main sources from which we obtained these data were libraries and archives in Germany, the U.K. and the U.S.A.[2]

Despite their different origin with respect to the source archive, each interview usually started with background questions regarding date and place of birth, and current age of the interviewee. Most interviewees then talked about their parents and other family members and their early lives in Germany, before re-telling their experiences of the pogrom on the night of the 9th to the 10th November 1938. For many this date was a turning point after which they knew that they would have to leave Germany soon. All of our interviewees left between the pogrom and the outbreak of WWII on September 1st, 1939. The testimonies usually proceeded with discussing the interviewees' process of emigration, their arrival in the country of destination, and the subsequent years.

Data generation. All interviews were audio-recorded. To overcome the differences in the quality of the recordings due to varying interview settings, all interviews were normalized through a standardized procedure involving the adjustment of loudness levels as well as noise and hiss reduction. The normalization procedure was done in *Adobe Audition* versions 3.0 and 6.0.During the normalization process, interviews were first converted to 22,050 Hz mono 16-bit. Next, their loudness levels were adjusted such that global loudness was kept within the 6dB boundary, with peaks not exceeding 0dB. This involved reducing the loudness of undesirable elements such as loud coughs, laughs, and etcetera. By means of the noise reduction feature in

[2] The data were obtained from the following institutions and archives: *Werkstatt der Erinnerung* (Hamburg, Germany), *Alte Synagoge Essen* (Germany), Prof. Manfred Brusten (Wuppertal, Germany), *USC Shoah Foundation Institute* (Los Angeles, U.S.A.), *Fortunoff Video Archive for Holocaust Testimonies* (Yale University Library, U.S.A.), *Tauber Holocaust Library and Education Program* (San Francisco, U.S.A.), *United States Holocaust Memorial Museum* (Washington, D.C., U.S.A.), and *the Association of Jewish Refugees* (London, U.K.).

Adobe Audition, in most cases the background noise could be successfully filtered out. In some cases, the frequencies were adjusted slightly to those most vital to speech (around 3 kHz) to accentuate the voice. For some recordings, hiss reduction was applied. This resulted in normalized sound files with an average sound level of -9dB to 12dB with all peaks around -3dB. Consequently, all sound beneath -20dB should be silence. Besides the previous normalization step, all fragments included in the rating experiment were adjusted such that the volume was equal.

For the fluency analyses, which were conducted with the *Praat* software package (version 5.3.67; Boersma & Weenink, 2014), the data had to be further processed. From each interview we extracted four one-minute fragments taken from minute 5, 10, 15, and 20 of each interview. This was to avoid bias, in case of tapping into emotional passages during someone's interview. In addition, we were careful to select one-minute passages where the interviewer did not interrupt the interviewee. We also removed pauses at the beginning and the end of each one-minute fragment before merging them into four-minute audio files for each interviewee.

The merged four-minute files were analyzed in *Praat* using the *Praat Script Syllable Nuclei v2* by De Jong and Wempe (2009) for the detection of syllable nuclei and the automatic measurement of speech rate. The threshold for empty pauses was set at 250ms given recent findings by De Jong and Bosker (2013). The minimum dip between peaks was set at -3 dB and for silences at -25 dB. The script outputs the following information: number of syllables, number of pauses, duration (in seconds), and phonation time (in seconds; excludes silent pauses), speech rate (number of syllables divided by duration), articulation rate (number of syllables divided by phonation time), and average speaking duration (ASD; speaking time divided by number of syllables). These measures will be explained in more detail in the subsequent section.

Fluency measures. Following Skehan's (2003) classification of *breakdown, speed* and *repair* fluency and in consideration of previous suggestions made by Bosker et al. (2013) and De Jong (2013) we generally used speaking time (measured in seconds) in the denominator, i.e. the actual time of speaking without pauses. By excluding pauses above 250ms from the speaking time we avoided a situation in which the different dimensions of fluency confounded, in particular measures of speed and repair with measures of breakdown. In accordance with prior findings on the relatedness of specific fluency measures with (measures of) L2 proficiency we selected one measure per breakdown, speed, and repair fluency. For an overview and the calculation of the measures, see Table 2.

Analyses. All analyses were run in *R* version 3.1.1. Fluency measures were outputted as csv files by *Praat* and loaded into *R*. Since each interviewee received multiple fluency scores, we conducted linear mixed-effects regression modeling, using the *R*-package *lme4* (version 1.1-6), with interviewee as a random-effect factor to take the structural variation linked to each interviewee into account (Winter, 2013). We assessed if random intercepts and random slopes were necessary by means of comparing the Akaike Information Criterion (AIC; Akaike, 1974). The AIC offsets the complexity of the model to the goodness of fit. An AIC difference of at least 2 (with the more complex model having a lower AIC) indicates that the higher complexity of the more complex model is warranted (Wieling, Nerbonne, & Baayen, 2011). AIC is related to the evidence ratio which expresses the relative probability that the model with the lowest AIC is more likely to provide a more precise model of the data (Blankevoort et al., 2013). Taking into account both random intercepts and slopes prevents being anti-conservative (i.e. reporting too high p-values; Baayen, 2008;Baayen, Davidson, & Bates, 2008).

To perform mixed-effects regression modeling on the different complexity scores, we first z-transformed all fluency measures. Using the reshape package (version 0.8.5) in *R*, we merged the scores into a fluency score, which was again z-transformed. The fluency score which comprises the three measures is hence a disfluency score: the higher the score (i.e., the more silent pauses, the longer the syllable duration and the more repairs), the more disfluent the speaker with respect to breakdown, speed, and repair fluency.

Results

The total number of observations in the dataset included in the analyses was 495 based on 99 interviewees. Table 2 shows the means and standard deviations for each individual fluency measure before *z*-transformation.

[Table 1]

Three interviewees had to be discarded due to insufficient sound quality of the audio recordings. Table 3 shows the coefficients and associated statistics of the fixed-effect factors and covariates of the final mixed-effects regression model obtained by using our exploratory analysis (the explained variance of the complete model including all random intercepts and slopes was: 5%; the fixed-effect predictors on their own accounted for: 3%).

[Table 3]

The model shows a significant interaction between type of measure and age at interview, which indicates that older interviewees produced significantly more silent pauses ($\beta = 0.292$, $t = 2.995$, $p < .01$) and repairs ($\beta = 0.199$, $t = 2.041$, $p < .05$). The relationship between age at interview and the scores for the different types of fluency is illustrated in Figure 1. The other potentially confounding variables (i.e. age of onset, age at interview, length of residence, and continued L1 exposure) did not reach significance independently or in interaction with any other variable and were therefore not included in the model.

[Figure 1]

Table 4 gives an overview of how log Likelihood and AIC values changed with the inclusion of the significant fixed-effect predictors while keeping the random-effects structure constant by including only random intercepts for interviewees (see Wieling et al., 2011). The baseline model only consisted of the random intercept for interviewee. The subsequent model (including the interaction between type of fluency and age at interview) was compared to the baseline model. The inclusion of the interaction between age at interview and fluency measure is warranted, given that its addition results in an AIC decrease of at least 2 each time.

[Table 4]

Model criticism revealed that the distribution of residuals was more or less normal, as illustrated in Figure 2.

[Figure 2]

Discussion

In this study we investigated temporal measures of fluency which have been said to cover three dimensions of utterance fluency, breakdown, speed, and repair, as proposed by (Skehan, 2003). For each of these dimensions we selected representative measures that have been previously shown to predict L2 proficiency and were therefore considered reliable measures (of L2 fluency). We then set out to test the effect of age of onset for the acquisition of L2 fluency, in addition to other variables such as continued L1 exposure or age at interview. Mixed-effects regression modeling revealed a significant interaction for type of measure and interview age. An older age at interview was found to predict a greater mean number of silent pauses and repairs.

However, the best model predicted only 5% of the variance of the data, suggesting that much remains unexplained. None of the other variables such as age of onset, continued L1 exposure, or level of education played a role.

Effects on temporal aspects of fluency. Unlike in the studies reviewed above (Derwing & Munro, 2013; Guion et al., 2000; Mackay & Flege, 2004; Trofimovich & Baker, 2006), we found no age of onset effect on any of the temporal measures considered here. This suggests that after having spent more than half of their lives in the L2 environment, our interviewees' delayed age of onset (AO) for learning the L2 left no traces on any of the three fluency dimensions. However, given the range of AO's we looked at, i.e. from 7 to 17 years, our findings are in line with Trofimovich & Baker's (2006) study of Italian immigrants to Canada. They had found an age effect, but only for immigrants who first started learning English in their late twenties. On the other hand, maximum length of residence (LoR) for the participants in their study was 10 years. Our interviewees had lived in the L2 environment well beyond 10 years by which time AO effects on fluency might disappear even for learners with AOs in their late 20ies and beyond. It is up to future research to test whether AO effects would be observable in L2 speakers with AOs beyond 17 but with extensive LoRs.

We found an interaction effect of fluency measure and age at interview indicating that breakdown and repair fluency suffer from (cognitive) aging, i.e. the older the interviewees at the time of the interview, the more silent pauses and repairs they produced. This is an indication of an aging effect as opposed to an age of onset effect (Stevens, 2006). A number of studies investigated disfluencies from the perspective of language and aging in monolinguals (for a review see Burke & Shafto, 2008). Based on picture-descriptions they found older adults to be generally more disfluent than younger adults. But findings are not consistent with regard to age differences across specific types of disfluency (Bortfeld, Leon, Bloom, Schober, & Brennan, 2001; Heller & Dobbs, 1993; Kemper, Rash, Kynette, & Norman, 1990; Schmitter-Edgecombe, Vesneski, & Jones, 2000 as cited in Burke & Shafto, 2008).

There are different positions with respect to explaining age differences in fluency. The existent evidence primarily suggests a cognitive decline resulting in an increase in lexical and/or phonological retrieval deficits in old age (Burke & Shafto, 2004). Alternatively, Ramscar, Hendrix, Shaoul, Milin, and Baayen (2014) declare cognitive decline a myth and attribute observed delays produced by older participants in lexical decision tasks to learning and the resulting increase in memory search demands rather than cognitive decline. As for our data, we did not observe an age at interview effect on any of the lexical measures, which were applied to the same interview data. This does not suggest greater lexical knowledge of interviewees with an older age at interview. Thus, cognitive decline appears to be a more plausible explanation, even though the question remains whether it concerns lexical and/or phonological retrieval deficits.

The aging effect only concerns the number of silent pauses per speaking time in seconds and the number of repairs per speaking time in seconds. Even though not significant, the *t*-value in Table 3 and the illustration in Figure 1 show that the effect of age at interview on syllable duration is reverse. Unlike De Jong, Steinel, et al. (2013) we found that number of silent pauses and the number of repairs correlated significantly negatively with syllable duration ($r = -.59$, $p < .0001$). This is interesting in two respects. First, it suggests that the more silent pauses and repairs occur in an interviewee’s production, the shorter the syllables, i.e. the higher the speed (when silent pauses are excluded). The increase of silent pauses and repairs might imply that the age at interview effect is primarily related to retrieval problems, although more precise measurements such as the location of pauses (possibly) in combination with speech errors would be necessary to give a better picture on what kind of retrieval problems interviewees are dealing with. Nevertheless, once retrieval problems are overcome, words can be produced at a similar pace independently of age at interview. Keeping Levelt's model in mind, this suggests that the deficit is more likely located at the stage of the articulator. Second, from a methodological point of view the inverse relationship (mean number of silent pauses and mean number of repairs correlate positively at $r = .46$, $p < .0001$) may also be an indication that syllable duration, representing the speed dimension according to Skehan's three-way distinction, does indeed captures another aspect of fluency.

Limitations. To begin with, a first limitation concerns the statistical analyses regarding the explained variance and the correlations between the three measures. The statistical model predicts little of the observed variance between interviewees and there is a lack of variance within interviewees. Not surprisingly, the explained variance of the fixed-effects regression model was similar to the comparable fixed-effect multiple regression model. In addition, we showed that the three fluency measures correlated. In this case an alternative would have been to build a composite value for all three dimensions of fluency by means of principal component analysis and to perform a traditional multiple regression analysis. However, given the theoretical discrimination between breakdown, speed, and repair and considering that syllable duration behaved in the opposite way as compared to number of silent pauses and number of repairs we wanted to keep the three measures apart.

Another limitation concerns the collinearity between length of residence and age at interview. These two independent variables correlated highly ($r = .9$). Upon suggestion by Wurm & Fisicaro (2014) we refrained however from residualizing in order to decorrelate the two variables. As Tabachnick and Fidell (2007) suggested, it is an option to ignore concerns of collinearity as long as the goal is to simply maximize the explained *variance*. During model building length of residence was not found significant in combination with age at interview and was therefore excluded.

A third limitation regards the amount of data based on which the measurements were generated. For each interviewee we only took four minutes of speech. Longer fragments would have been preferable, but the data are interviews, which are dialogic in nature. Even though interviewees went through phases of monologue, giving extensive answers to certain questions, the testimonies are marked by interruptions and exchanges between the interviewer and the interviewee.

Finally, several recent studies by Forsberg and colleagues assessed the use of multiword units which have been suggested as devices of fluency (e.g., Pawley & Syder, 1983). These studies found that late-onset learners with comparable lengths of residence (as those in the present study) did not show nativelike performance. In the present study we focused exclusively on the use of temporal measures to capture fluency. Investigations of multiword units at the ultimate stages of L2 development for the assessment of fluency seem to be a worthwhile area for future research.

Conclusion. Despite these limitations we conclude that our interviewees are no longer affected by the usual factors of L2 acquisition, i.e. age of onset or continued L1 exposure, at the productive level where temporal aspects of fluency are being assessed. These results confirm some of our interviewees' judgments regarding fluency in the broad sense as presented initially. It seems indeed that our interviewees had quickly become fluent (and proficient) L2 speakers of

English, but like any language user, our interviewees are also susceptible to an age-related cognitive decline for this domain of (second) language production.

References:

Abrahamsson, N., & Hyltenstam, K. (2009). Age of Onset and Nativelikeness in a Second Language: Listener Perception Versus Linguistic Scrutiny. *Language Learning*, *59*(2), 249–306.

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, *19*(6), 716–723.

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, *59*(4), 390–412.

Birdsong, D., & Molis, M. (2001). On the Evidence for Maturational Constraints in Second-Language Acquisition. *Journal of Memory and Language*, *44*(2), 235–249.

Blankevoort, C. G., Scherder, E. J. A., Wieling, M. B., Hortobágyi, T., Brouwer, W. H., Geuze, R. H., & van Heuvelen, M. J. G. (2013). Physical Predictors of Cognitive Performance in Healthy Older Adults: A Cross-Sectional Analysis. *PLoS ONE*, *8*(7), e70799.

Boersma, P., & Weenink, D. (2014). Praat: doing phonetics by computer [Computer program] (Version 5.3.76). Retrieved from http://www.praat.org/

Bortfeld, H., Leon, S. D., Bloom, J. E., Schober, M. F., & Brennan, S. E. (2001). Disfluency Rates in Conversation: Effects of Age, Relationship, Topic, Role, and Gender. *Language and Speech*, *44*(2), 123–147.

Bosker, H. R., Pinget, A. F., Quené, H., Sanders, T., & De Jong, N. H. (2013). What makes speech sound fluent? The contributions of pauses, speed and repairs. *Language Testing, 30*(2), 159-175.

Burke, D. M., & Shafto, M. A. (2008). Language and Aging. In F. I. M. Craik & T. A. Salthouse (Eds.), *The Handbook of Aging and Cognition* (pp. 373–443). New York, NY: Psychology Press.

Chambers, F. (1997). What do we mean by fluency? *System*, *25*(4), 535–544.

Cucchiarini, C., Strik, H., & Boves, L. (2000). Quantitative assessment of second language learners' fluency by means of automatic speech recognition technology. *The Journal of the Acoustical Society of America*, *107*(2), 989–999.

Cucchiarini, C., Strik, H., & Boves, L. (2002). Quantitative assessment of second language learners' fluency: Comparisons between read and spontaneous speech. *The Journal of the Acoustical Society of America*, *111*(6), 2862–2873.

De Jong, N. H. (2013). *Unraveling measures of fluency*. Presented at the EUROSLA23, Amsterdam. Retrieved from http://prezi.com/qpi0fdpzxlqb/how-do-i-choose-which-measure-of-fluency-to-use/

De Jong, N. H., & Bosker, H. R. (2013). Choosing a threshold for silent pauses to measure second language fluency. In *Proceedings of Disfluency in Spontaneous Speech, DiSS* (pp. 17–20). Stockholm: Universitetsservice US-AB.

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 behavior. *Applied Psycholinguistics*, *34*, 1–21.

De Jong, N. H., Steinel, M. P., Florijn, A., Schoonen, R., & Hulstijn, J. H. (2013). Linguistic skills and speaking fluency in a second language. *Applied Psycholinguistics*, *05*(34), 893–916.

De Jong, N. H., & Wempe, T. (2009). Praat script to detect syllable nuclei and measure speech rate automatically. *Behavior Research Methods*, *41*(2), 385–390.

Derwing, T. M., & Munro, M. J. (2013). The Development of L2 Oral Language Skills in Two L1 Groups: A 7-Year Study. *Language Learning*, *63*(2), 163–185.

Derwing, T. M., Munro, M. J., Thomson, R. I., & Rossiter, M. J. (2009). The relationship between L1 fluency and L2 fluency development. *Studies in Second Language Acquisition*, *31*(4), 533–557.

Derwing, T. M., Rossiter, M. J., Munro, M. J., & Thomson, R. I. (2004). Second Language Fluency: Judgments on Different Tasks. *Language Learning*, *54*(4), 655–679.

Elsendoorn, B. A. G. (1984). *Tolerances of durational properties in British English vowels*. (Doctoral Dissertation).

Fillmore, C. J. (1979). On fluency. In *Individual differences in language ability and language behavior* (pp. 85–102). New York: Academic Press.

Flege, J. E. (1979). *Phonetic interference in second language acquisition.* (Unpublished doctoral dissertation). Department of Linguistics, Indiana University, Bloomington.

Forsberg Lundell, F., Bartning, I., Engel, H., Gudmundson, A., Hancok, V., & Lindqvist, C. (2013). Beyond advanced stages in high-level spoken L2 French. *Journal of French Language Studies*, *24*(2), 255–280.

Gelderen, A. van. (1994). Prediction of global ratings of fluency and delivery in narrative discourse by linguistic and phonetic measures- oral performances of students aged 11-12 years. *Language Testing*, *11*(3), 291–319.

Ginther, A., Dimova, S., & Yang, R. (2010). Conceptual and empirical relationships between temporal measures of fluency and oral English proficiency with implications for automated scoring. *Language Testing*, *27*(3), 379–399.

Götz, S. (2013). *Fluency in Native and Nonnative English Speech*. John Benjamins Publishing.

Granena, G., & Long, M. H. (2013). Age of onset, length of residence, language aptitude, and ultimate L2 attainment in three linguistic domains. *Second Language Research*, *29*(3), 311–343.

Guion, S. G., Flege, J. E., Liu, S. H., & Yeni-Komshian, G. H. (2000). Age of learning effects on the duration of sentences produced in a second language. *Applied Psycholinguistics*, *21*(02), 205–228.

Heller, R. B., & Dobbs, A. R. (1993). Age differences in word finding in discourse and nondiscourse situations. *Psychology and Aging*, *8*(3), 443–450.

Hellman, A. B. (2011). Vocabulary size and depth of word knowledge in adult-onset second language acquisition. *International Journal of Applied Linguistics*, *21*(2), 162–182.

Housen, A., Kuiken, F., & Vedder, I. (2012). *Dimensions of L2 Performance and Proficiency: Complexity, Accuracy and Fluency in Sla*. Amsterdam: John Benjamins Publishing.

Iwashita, N. (2010). Features of oral proficiency in task performance by EFL and JFL learners. In *Selected Proceedings of the 2008 Second Language Research Forum* (pp. 32–47).

Iwashita, N., Brown, A., McNamara, T. i., & O'Hagan, S. (2008a). Assessed Levels of Second Language Speaking Proficiency: How Distinct? *Applied Linguistics*, *29*(1), 24–49.

Iwashita, N., Brown, A., McNamara, T., & O'Hagan, S. (2008b, March 1). Assessed Levels of Second Language Speaking Proficiency: How Distinct? *Applied Linguistics*, *29*(1), 24 – 49.

Johnson, J. S., & Newport, E. L. (1989). Critical period effects in second language learning: The influence of maturational state on the acquisition of English as a second language. *Cognitive Psychology*, *21*(1), 60–99.

Johnson, J. S., & Newport, E. L. (1991). Critical period effects on universal properties of language: The status of subjacency in the acquisition of a second language* 1. *Cognition*, *39*(3), 215–258.

Kemper, S., Rash, S., Kynette, D., & Norman, S. (1990). Telling stories: The structure of adults' narratives. *European Journal of Cognitive Psychology*, *2*(3), 205–228.

Kormos, J., & Dénes. (2004). Exploring measures and perceptions of fluency in the speech of second language learners. *System*, *32*(2), 145–164.

Lennon, P. (1990). Investigating Fluency in EFL: A Quantitative Approach. *Language Learning*, *40*(3), 387–417.

Lennon, P. (2000). The lexical element in spoken second language fluency. In H. Riggenbach (Ed.), *Perspectives on fluency* (pp. 25–42). Ann Arbor, MI: University of Michigan Press.

Levelt, W. J. M. (1989). *Speaking: from intention to articulation*. Cambridge, MA: MIT Press.

MacKay, D. G. (1987). *The Organization of Perception and Action: A Theory for Language and Other Cognitive Skills*. New York, NY: Springer-Verlag New York, Inc.

Mackay, I. R. A., & Flege, J. E. (2004). Effects of the age of second language learning on the duration of first and second language sentences: The role of suppression. *Applied Psycholinguistics*, *25*(03), 373–396.

Mizera, G. J. (2006). Working memory and L2 Oral Fluency. Retrieved November 6, 2012, from http://d-scholarship.pitt.edu/7655/

Munro, M. J., & Derwing, T. M. (1995). Foreign Accent, Comprehensibility, and Intelligibility in the Speech of Second Language Learners. *Language Learning*, *45*(1), 73–97.

Oyama, S. (1976). A sensitive period for the acquisition of a nonnative phonological system. *Journal of Psycholinguistic Research*, *5*(3), 261–283.

Pawley, A., & Syder, F. H. (1983). Two puzzles for linguistic theory: Nativelike selection and nativelike fluency. In J. C. Richards & R. Schmidt (Eds.), *Language and Communication* (pp. 191–225). London: Longman.

Piske, T., MacKay, I. R. A., & Flege, J. E. (2001). Factors affecting degree of foreign accent in an L2: a review. *Journal of Phonetics*, *29*(2), 191–215.

Ramscar, M., Hendrix, P., Shaoul, C., Milin, P., & Baayen, H. (2014). The Myth of Cognitive Decline: Non-Linear Dynamics of Lifelong Learning. *Topics in Cognitive Science*, *6*(1), 5–42.

Rehbein, J. (1987). On fluency in second language speech. In H. W. Dechert & M. Raupach (Eds.), *Psycholinguistic models of production* (Norwood, pp. 97–105). Ablex.

Riazantseva, A. (2001). Second language proficiency and pausing: A Study of Russian Speakers of English. *Studies in Second Language Acquisition*, *23*(04), 497–526.

Rossiter, M. J. (2009). Perceptions of L2 Fluency by Native and Non-native Speakers of English. *Canadian Modern Language Review/ La Revue Canadienne Des Langues Vivantes*, *65*(3), 395–412.

Schmid, M. S., & Dusseldorp, E. (2010). Quantitative analyses in a multivariate study of language attrition: the impact of extralinguistic factors. *Second Language Research*, *26*(1), 125 –160.

Schmidt, R. (1992). Psychological Mechanisms Underlying Second Language Fluency. *Studies in Second Language Acquisition*, *14*(04), 357–385.

Schmitter-Edgecombe, M., Vesneski, M., & Jones, D. W. R. (2000). Aging and Word-Finding: A Comparison of Spontaneous and Constrained Naming Tests. *Archives of Clinical Neuropsychology*, *15*(6), 479–493.

Segalowitz, N. (2010). *Cognitive Bases of Second Language Fluency*. Abingdon, Oxon: Routledge.

Skehan, P. (2003). Task-based instruction. *Language Teaching*, *36*(01), 1–14.

Stevens, G. (2006). The Age-Length-Onset Problem in Research on Second Language Acquisition Among Immigrants. *Language Learning*, *56*(4), 671–692.

Towell, R., Hawkins, R., & Bazergui, N. (1996). The Development of Fluency in Advanced Learners of French. *Applied Linguistics*, *17*(1), 84–119.

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*(01), 1–30.

Wieling, M., Nerbonne, J., & Baayen, R. H. (2011). Quantitative Social Dialectology: Explaining Linguistic Variation Geographically and Socially. *PLoS ONE*, *6*(9), e23613.

Winter, B. (2013). *Linear models and linear mixed effects models in R with linguistic applications.* Retrieved from http://arxiv.org/pdf/1308.5499.pdf

Wurm, L. H., & Fisicaro, S. A. (2014). What residualizing predictors in regression analyses does (and what it does not do). *Journal of Memory and Language*, *72*, 37–48.

Table 1. Overview independent variables (fluency)

	n	*M* (*SD*)	Range/Categories		
Age of onset	101	12.15 (2.67)	7-17		
Length of residence	100	61.33 (6.12)	41-73		
Age at interview	101	73.59 (6.97)	57-87		
L1 exposure	98	4.34 (1.44)	1-7		
L1 at work	80		Yes: 14	No: 66	
Gender			M: 42	F: 60	
Edu			Low: 11	Mid: 34	High: 46

Note: AO = age of onset, LoR = length of residence, AaI = age at interview, L1 Exp = continued L1 exposure, L1 at work = use of German at work, M = male, F = female, Edu = level of education.

Table 2. Overview of fluency measures

Aspect	**Acoustic measure**	**Calculation**	**Mean (SD)**
Breakdown	Mean number of silent pauses (silent pauses)	Number of silent pauses/spoken time (in seconds)	0.69 (0.26)
Speed	Mean syllable duration (syllable duration)	Spoken time (in seconds)/number of syllables	0.21 (0.03)
Repair	Mean number of repairs (repairs)	Log(repairs/spoken time (in seconds))	-2.68 (0.57)

Note: We show here all measures before z-transformation for the mixed-effect regression.

Table 3. Linear mixed effect model (fluency)

Fixed effect	**Estimate**	**Std. Error**	***t*-value**[1]	***p*-value**
Intercept [silent pauses]	0.006	0.098	0.060	.95
FluTypeSyllable duration	-0.009	0.138	-0.064	.95
FluTypeRepairs	-0.002	0.138	-0.014	.99
FluTypeSilentPauses:age at interview	0.292	0.098	2.995	< .01**
FluTtypeSyllable duration:age at interview	-0.149	0.098	-1.527	.13
FluTypeRepairs:age at interview	0.199	0.098	2.041	< .05*

Note: FluType = Fluency Type

Table 4. Goodness of fit of the fixed-effects structure (fluency)

Additional fixed effects	Log likelihood increase	– AIC decrease	Evidence ratio	Likelihood ratio test	Additional degrees of freedom
Random intercept only					
+ Age at interview:Fluency measure	7.54	5.08	> 1000	*p* < .01	3

Note: Each row specifies the significant increase in goodness of fit obtained by adding the current predictor to the model including all preceding predictors. AIC: Akaike Information Criterion.