# Attached filled pauses: Occurrences and durations

*Mária Gósy [1], Vered Silber-Varod [2]*
*[1]Hungarian Research Centre for Linguistics and ELTE University, Budapest, Hungary*
*[2]Open Media and Information Lab (OMILab), The Open University of Israel, Israel*

## *Abstract*

*Filled pauses may reveal speech planning or execution problems that result in various positional and temporal patterns in spontaneous utterances. The purpose of this study was to analyze the position of the vocalic FPs, with respect to an adjacent word, in terms of occurrences and their durations produced by young (mean age: 25 years) and elderly (mean age: 76 years) speakers of Hungarian (a total of 32 participants). Elderly speakers produced significantly less and longer vocalic FPs than young speakers did. Both the occurrences and durations were significantly influenced by position of FPs and by age. In this paper, we introduced the conception of a functional difference between FPs attached either to the preceding or to the following word. The findings indicated different ways of resolving speech planning or execution problems depending on age.*

## Introduction

The phenomenon identified by the term 'filled pause' (henceforward FP) has been known and studied for about 60 years (Maclay & Osgood, 1959; Shriberg, 2001; Fox Tree, 2002; Corley & Stewart, 2008, etc.). In Hungarian, [ø]-like or [ə]-like vowels are the most frequent types to fill pauses in spontaneous utterances (e.g., Gósy et al., 2014). FPs were found to have various functions. They provide extra time for the speaker to aid speech planning and execution, monitoring and repair, as well as they signal conversational turns or occurring as discourse markers, etc. (e.g., Smith & Clark, 1993; Fox Tree, 2002; Watanabe et al., 2008; Finlayson & Corley, 2012; Urizar & Samuel, 2014). There are, however, serious difficulties when one tries to identify, separate or categorize these functions (Cutler, 1988).

There is a specific property of FPs that they can occur either between two silent pauses or attached to (co-articulated with) a word. FPs may occur inserted before the first segment of the word (FPword position) or after the last segment of the word (wordFP position). Clark and Fox Tree (2002) termed these two positions as cliticization. There are a limited number of papers focusing on the positioning of FPs related to the neighboring words in the literature. Speakers were reported to attach a FP onto a previous word, but not onto a following in native (British) English speech (Clark & Fox Tree, 2002). However, FPs were more frequent between a lexical item and a silent pause than between two silent pauses in another study also in (British) English speech (Leeuw, 2007). Dutch and German speakers seemed to behave differently in positioning FPs; attached FPs were found to be common in Dutch while there was no difference in the positions of FPs in German (Leeuw, 2007). Silber-Varod and colleagues (2016) found that in Hebrew, attached FPs are more common than FPs between silent pauses, and enclitic FPs (wordFP) are more common than proclitic FPs (FPword). The articulation gesture of attaching a FP to a word is easy to perform and provides a kind of concealment of the FP (and the speaker's difficulty) since it is not flashy in these positions.

Findings about the occurrences and durations of FPs in various age groups seem to be controversial (e.g., Bortfeld et al., 2001; Searl et al., 2002). Some studies reported that elderly people used a larger number of FPs as opposed to young adults (Bortfeld et al., 2001; Roggia, 2012). Kemper (1992) found that old-old speakers (ages between 75 and 90) produced more FPs than young-old speakers (ages between 60 and 74) did. In contrast, other studies did not find such differences (Leeper Calcutta 1995; Bóna, 2014; Gósy et al., 2014). Bóna (2011) found that as soon as the topic of the narrative became more challenging, young subjects produced a higher number of FPs than old subjects did. Emotional stress seemed to influence elderly speakers' pausing more than those of young ones (Caruso et al., 1997).

The durational range of FPs is wide, and there are a great many factors that influence the measured values (average and contextual speech rate, thinking speed, difficulty of the topic to be discussed, etc.). The mean durations of FPs are reported to range from about 100 ms to about 750 ms or even longer (e.g., Shriberg, 2001; Clark & Fox Tree, 2002; Merlo & Barbosa, 2010; de Jong & Bosker, 2013). FPs' durations were shown to increase by age in some studies (e.g., Pindzola, 1990; Kemper, 1992) while others did not support significant differences between young and old speakers in this respect (e.g., Horton et al., 2011; Bóna, 2014; Gósy et al., 2014). Durations of FPs occurring between two silent pauses were significantly longer than those occurring between a lexical item and a silent pause in two middle-aged Hungarian-speaking speakers' spontaneous utterances (Gósy, 2015).

We are of the opinion that more clear patterns can be found on the (sometimes controversial) behavior of FPs if the factor of their immediate position is also considered. The main body of the present study addresses whether occurrences and durations of FPs show differences depending on age (young and old speakers) and positions with regards to adjacent words and silent pauses in Hungarian. We hypothesized that (i) the occurrences of FPs would show significantly different patterns depend-ing on age, (ii) the proportions of FPs in various positions would show significant differences, (iii) the durations of FPs would show significant differences depending on age, and (iv) the durations of FPs would show significant differences depending on their positions.

## Methodology

Thirty-two spontaneous narratives produced by native Hungarian speakers (half of them were females) were randomly selected (with the exception of age and gender criteria) from the BEA Hungarian speech database (Gósy, 2012). Two distinct age groups were formed: (i) young speakers (aged between 22 and 28 years; mean age 25 years) and (ii) old speakers (aged between 70 and 80 years; mean age: 76 years). The participants were asked to speak about their life and about their opinions on topics of current interest provided by the interviewer (who was the same person across all recordings). The mean speech rate was 4.4 syllables/s in young speakers while 3.8 syllables/s in old speakers.

Recordings were made in the same sound-attenuated room, under identical technical conditions using an Audiotechnica AT4040 cardioid condenser microphone connected directly to a computer using GoldWave to record samples at 44.1 kHz, 16 bits, monaurally. For the present study, more than 4.5 hours of speech samples from the database were used. The duration of recording per subject was 8.5 minutes (std. dev. = 0.3 minutes).

The speech material was manually annotated focusing on vocalic FPs (variants of the [ø] vowel and the [ə] neutral vowel). FPs were marked by öö while silent pauses were marked by SIL in annotations. Silent periods may contain also breath noise (Trouvain et al., 2016). The positions of vocalic FPs were also coded as between silent pauses or silence on one side and lexeme on the other. Transcription was done in Praat (Boersma & Weenink, 2016). Occurrences of all FPs were analyzed according to the three possible positions. There are two instances where FPs are attached to the lexical items. FP can be attached to the first segment of the word (this is the FPword position) after a silent pause. FP can be attached to the last segment of the word (this is the wordFP position) and is followed by a silent pause. The third position occurs when FP is surrounded by silent pauses on both sides (this is the silFPsil position). (The occurrences of FPs surrounded by two words were extremely rare, thus, they were excluded from the analysis.) Examples: (i) *le akartam fényképezni a* SIL *ööhegyeket* ('I wanted to take pictures /of/ SIL öömountains'); (ii) *tehátöö* SIL *furcsa helyzet van* ('wellöö SIL there is a peculiar situation'); *probléma hogy ez* silFPsil *nehéz feladat* ('the problem is that this /is/ silFPsil /a/ difficult task').

A total of 1,068 FPs (284 of them were produced by the old and 784 by the young speakers) were found in the speech material. The total number of silent pauses were 1271, out of 904 were found in young while 367 in old speakers. The items of silFPsil type consisted of 240 silent pauses in young and 166 silent pauses in old speakers. Figure 1 shows spectrograms of FPs in the attached positions.

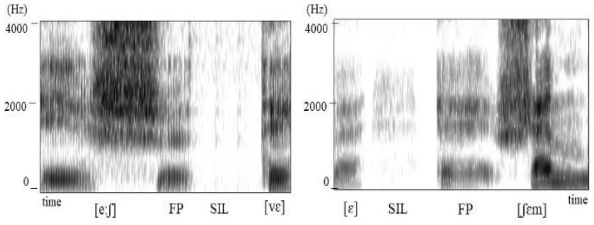


*Figure 1. Speech fragments where FP is followed (left) and preceded (right) by a silent pause (containing some breath) and is attached to the first consonant [ʃ] of the following word* sem *'neither' (right) and to the last consonant [ʃ] of the preceding word* és *'and' (left).*

The duration was measured as the interval (i) between the onset and offset of the second formants of the vocalic FP occurring between silent pauses, and (ii) between the onset/offset of the second formant of the vocalic FP and the onset/offset of the preceding and following segment based on traditional criteria. Durations were extracted using a specific Praat script. All inter-lexical pauses (Zellner, 1994) were considered. The shortest silent pause in the vicinity of FPs was 40 ms. Prolongations and outlier data were excluded from (further) analysis.

To test statistical significance, MANOVA was performed on durations of vocalic FPs as dependent factors. As fixed effects, we entered 'age group', and 'position' into the model. Chi-Square and Mann–Whitney U tests were performed to analyze occurrences of FPs. In all cases, the confidence level was set at the conventional 95%.

## Results

### *Occurrences*

Considering all FPs, we found 5.7 incidents per minute in young speakers while 2.1 incidents per mi-

nute in old speakers. Statistical analysis revealed significant differences depending on both 'position' and 'age' (for position: Chi-Square = 131.511; $p < 0.001$; for age: Chi-Square = 234.082; $p < 0.001$). Young speakers produced the incidents of wordFP type in 3.1 per minute, while they produced the incidents of FPword type less frequently (1.8 incidents per minute). The incidents of the type silFPsil occurred the least frequently in their speech samples (0.9 incidents per minute).

Old speakers produced the incidents of FPword type in 0.8 per minute while there were no significant differences in occurrences of the incidents of wordFP and silFPsil types in their case (0.6 incidents per minute in both cases: Mann–Whitney U test: $Z = 0.957$, $p > 0.05$).

The distribution of FPs according to position, within an age group, showed significant differences in both young (Chi-Square=179.429; $p < 0.001$) and elderly speakers (Chi-Square=7.232; $p < 0.027$). Figure 2 demonstrates the different ratios of occurrences according to FP types expressed in percentages in both age groups.

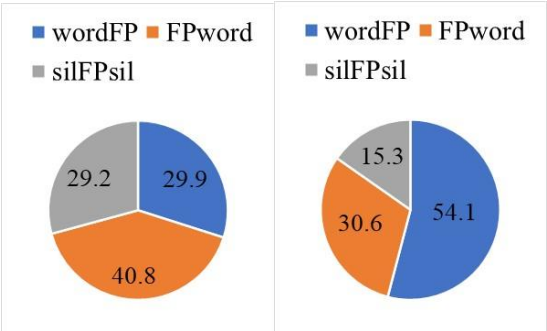


*Figure 2. Occurrences of FPs (%) according to the three position types (left side: old speakers, right side: young speakers).*

### Durations

Young speakers produced significantly shorter FPs compared to old speakers (mean value: 411 ms; standard deviation /std. dev./: 309 ms) versus mean value: 479 ms; std. dev.: 207 ms, respectively). Statistical analysis revealed significant differences in the durations of FPs depending on both factors of 'age' ($F(1, 1067) = 4.181$; $p = 0.041$; *partial* $\eta^2 = 0.004$) and 'position' ($F(1, 1067) = 37.905$; $p = 0.001$; *partial* $\eta^2 = 0.067$). In both groups, the shortest FPs were produced in FPword positions (mean value of young speakers: 341 ms, std. dev.: 353 ms; mean value of old speakers: 408 ms, std. dev.: 175 ms) while the longest ones were produced in silFPsil positions (mean value of young speakers: 606 ms, std. dev.: 286 ms; mean value of old speakers: 575 ms, std. dev.: 178 ms). The durations of FPs in wordFP positions were in between the other two types, both in young as well as in old speakers (mean value of young speakers: 395 ms, std. dev.: 264 ms; mean value of old speakers: 481 ms, std. dev.: 237 ms). Post hoc Tukey test revealed significant differences in durations of FPs, depending on positions in all comparisons ($p < 0.005$). The interaction of age and position was not statistically significant ($F(1, 1067) = 2.968$; $p = 0.052$; *partial* $\eta^2 = 0.006$). The durations of FPs in various positions seem to form the same patterns irrespective of age (Figure 3).

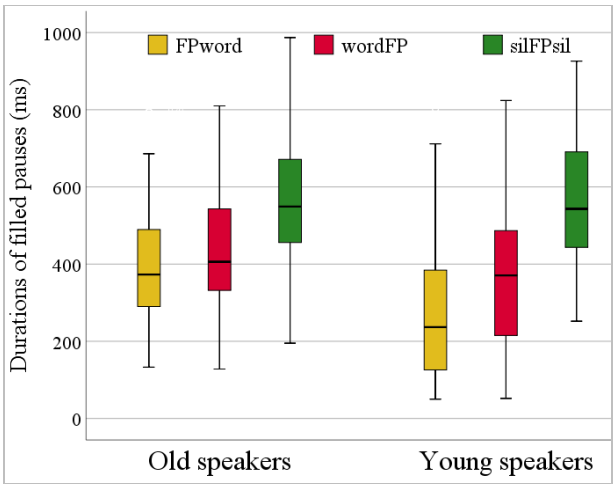


*Figure 3. Durations of FPs in the three positions produced by young and old speakers (quartiles, medians).*

The durations of the silent pauses produced by both young and old speakers preceding FPs were significantly different according to 'position' but not to 'age' ($F(2, 1067) = 249.78$, $p = 0.001$; *partial* $\eta^2 = 0.320$; $F(1, 1067) = 3.470$; $p = 0.063$; *partial* $\eta^2 = 0.002$, respectively). Post hoc Tukey tests revealed significant differences in all cases ($p < 0.05$). The interaction of 'position' and 'age' was not significant ($F(2, 1067) = 1.090$, $p = 0.337$). The durations of the silent pauses produced by both young and old speakers following FPs were significantly different according to both 'position' and 'age' ($F(2, 1067) = 79.770$, $p = 0.001$; *partial* $\eta^2 = 0.131$; $F(1, 1067) = 13.534$; $p = 0.001$; *partial* $\eta^2 = 0.02$, respectively). Post hoc Tukey tests revealed significant differences in all cases ($p < 0.05$). The interaction of position and age was significant ($F(2, 1067) = 4.201$, $p = 0.008$, *partial* $\eta^2 = 0.008$).

Silent pauses were shorter before FPword position than after wordFP position in young speakers; however, the opposite could be found in old speakers (Table 1). In silFPsil positions, the first silent pauses were significantly longer than the second ones only in old speakers ($F(1, 165) = 13.922$, $p = 0.001$; *partial* $\eta^2 = 0.078$; $F(1, 239) = 0.934$, $p = 0.335$; *partial* $\eta^2 = 0.004$, respectively).

*Table 1. Mean durations of silent pauses according to positions (given in ms, std. dev. values are in brackets).*

| Age group | silFPword | wordFPsil | $sil_1FP$ | $FPsil_2$ |
|---|---|---|---|---|
| Young | 484 (384) | 537 (481) | 725 (478) | 665 (549) |
| Old | 438 (295) | 386 (327) | 631 (505) | 383 (330) |

## Conclusions

The goal of this study was to obtain information on occurrence of positions and durations of vocalic FPs produced by young and old Hungarian-speaking speakers. Young speakers produced close to twice as many FPs in their utterances than old speakers did. Our first hypothesis was confirmed. These findings add to the controversial data of the literature mentioned earlier (e.g., Bortfeld et al., 2001; Leeper & Culatta, 1995; Searl et al., 2002; Gósy et al., 2014). The attached FPs were all longer and less frequent in old speakers while they were shorter and more frequent in young speakers. However, practically no difference was found in the durations of FPs occurring between two silent pauses between young and old speakers. This finding suggests that these FPs might function as discourse markers with distinct traits from attached FPs. As such, both the occurrences and durations of FPs showed significantly different patterns, depending on age.

Our explanation for the findings comprises two aspects: Speech planning differences between young and old speakers as well as different speaking routines. The different speech rates of the young an old speakers may also contribute to temporal differences. Old people are supposed to activate their thoughts at a slower speed compared to young speakers. Thus, they might not often need extra time for selection of thoughts. The simultaneity of activation and selection of thoughts together with old speakers' simpler grammatical structures (see Horton et al., 2011) seem to be more transparent and more easily managed. Young speakers are supposed to activate numerous thoughts at the same time at a high speed that require continuous activation and selection of thoughts followed by transforming them into grammatical forms. These tasks require extra time. Obviously, old speakers have more routine in verbal communication including well-tried strategies as opposed to young speakers. We assumed that the proportions and durations of FPs in various positions would show significant differences which was confirmed. In FPword position, the speaker tries to solve the problem during the silent pause that precedes FP; however, this amount of time is not enough, therefore, the speaker starts producing a FP which increases the necessary time to continue (Maschler, 2001). The more frequent occurrence of this phenomenon in old speakers reflects that they are not able to solve the problem during speech planning, consequently they are in need for more time. In wordFP position, the speaker anticipates some problem with the continuation during the (last) word production. Young speakers' monitoring works better and faster than those of old speakers that provides an early identification of the problem. To gain extra time, the speaker lengthens the word by means of coarticulating a FP followed by a silent pause. This strategy is similar to that of segment prolongation phenomenon. If the silent pause+FP combination is not sufficient for problem solving, another silent pause is added (silFPsil position).

We suggest the conception of a functional difference between the FPword and wordFP positions. FPword position signals a speech planning problem while wordFP position signals the repair that is happens during that time. Speech planning problem means re-selection and/or re-organization of thoughts that needs longer time to perform as opposed to repairing obvious errors. This interpretation is supported by durations since FPs in the FPword positions were longer than those in the wordFP positions. FPs surrounded by two silent pauses may signal both the planning problem (first silent pause and FP) as well as the momentary inability to repair it (FP and second silent pause) or some other processing strategy. The longer durations of FPs surrounded by silent pauses compared them to those of attached FP types support this interpretation. Thus, we arrived at the conclusion that the FP position has a functional definiteness.

The data highlight some temporal equalization in the case of the attached FPs in both age groups. The mean durations of the silent pause and FP combinations are around 850 ms, indicating the inner control over the speech planning disharmony by the speaker. The silFPsil combinations show the tendency that the silent pauses following FPs are shorter than those preceding FPs. This tendency can be explained by the fact that FPs signal the (near-)resolution of the problem-solving process.

Our findings call attention to the interrelations of immediate positions and durations of FPs in spontaneous utterances that provide a better understanding of the surface effects of the speakers' speech planning difficulties.

## Acknowledgements

This research has been implemented with the support provided by the Ministry of Innovation and Technology of Hungary from the National Research, Development, and Innovation Fund, financed under the ELTE TKP2020-IKA-06 funding scheme.

## References


Boersma, P. & D. Weenink. 2021. *Praat: doing phonetics by computer* [Computer program]. Version 6.1.40, retrieved 27 February 2021 from http://www.praat.org/

Bóna, J. 2011. Disfluencies in the spontaneous speech of various age groups: Data from Hungarian. *Govor* 28: 95–115.

Bóna, J. 2014. Temporal characteristics of speech: The effect of age and speech style. *Journal of the Acoustical Society of America* 136(2), EL116–121. https://doi.org/ 10.1121/1.4885482

Bortfeld, H., S. D. Leon, J. E. Bloom, M. F. Schober, & S. E. Brennan. 2001. Disfluency rates in conversation: Effects of age, relationship, topic, role, and gender. *Language and Speech* 44(2): 123–147.

Caruso, A. J., T. M. McClowry, & M. Ludo. 1997. Age-related effects on speech fluency. *Seminars in Speech and Language* 18: 171–180. https://doi.org/ 10.1055/s-2008-1064071

Clark, H. H. & J. E. Fox Tree, 2002. Using *uh* and *um* in spontaneous speaking. *Cognition* 84(1), 73–111. https://doi.org/ 10.1016/s0010-0277(02)00017-3

Corley, M. & O. W. Stewart. 2008. Hesitation disfluencies in spontaneous speech: the meaning of *um*. *Language and Linguistics Compass* 2(4): 589–602. https://doi.org/10.1111/j.1749-818x.2008.00068.x

Cutler, A. 1988. The perfect speech error. In: L. M. Hyman & C. N. Li (eds.): *Language, speech, and mind: Studies in honor of Victoria A. Fromkin*. Routledge, London, 209–223.

Finlayson, I. R. & M. Corley. 2012. Disfluency in dialogue: An intentional signal from the speaker? *Psychonomic Bulletin & Review* 19(5): 921–928. https://doi.org/ 10.3758/s13423-012-0279-x

Fox Tree, J. E. 2002. Interpreting pauses and ums at turn exchanges. *Discourse Processes* 34(1): 37–55. https://doi.org/ 10.1207/s15326950dp3401_2

Gósy, M. 2012. BEA – A multifunctional Hungarian spoken language database. *The Phonetician* 105/106: 50–61. http://www.isphs.org/Phonetician/Phonetician-105-106.pdf

Gósy, M. 2015. Beszédtervezési diszharmónia és a kitöltött szünetek összefüggései. [Interrelations of speech planning disharmony and FPs] *Magyar Nyelvőr* 139/4: 436–449.

Gósy, M., J. Bóna, A. Beke, & V. Horváth. 2014. Phonetic characteristics of filled pauses: the effects of speakers' age. *10th ISSP*, Cologne, 150–153. http://real.mtak.hu/ id/eprint/15491

Horton, W. S., D. H. Spieler, & E. Shriberg. 2011. A corpus analysis of patterns of age-related change in conversational speech. *Psychology of Aging* 25(3): 708–713. https://doi.org/ 10.1037/a0019424

Jong, N. H. de & H. R. Bosker, 2013. Choosing a threshold for silent pauses to measure second language fluency. In Eklund, R. (ed.): *Proceedings of DiSS 2013: The 6th Workshop on Disfluency in Spontaneous Speech*. KTH Royal Institute of Technology, Stockholm, Sweden: TMH-QPSR Vol. 54(1), 17–20.

Kemper, S. 1992. Language and aging. In: Craik FIM, Salthouse T. A. (eds.) *The Handbook of Aging and Cognition*. Hillsdale, NJ, England: Lawrence Erlbaum Associates, 213–270. https://doi.org/10.1080/01924780903295796

Leeper, L. H. & R. Culatta. 1995. Speech fluency: Effect of age, gender and context. *Folia Phoniatrica et Logopedia* 47: 1–14. https://doi.org/ 10.1159/000266 337

Leeuw de, E. 2007. Hesitation markers in English, German, and Dutch. *Journal of Germanic Linguistics* 19(2): 85–114. https://doi.org/10.1017/s1470542707000049

Maclay, H. & C. Osgood. 1959. Hesitation phenomena in spontaneous English speech. *Word* 15(1): 19–44. https://doi.org/ 10.1080/00437956.1959.11659682

Maschler, Y. 2009. *Metalanguage in interaction: Hebrew discourse markers*. Amsterdam: John Benjamins. https://doi.org/10.1075/pbns.181

Merlo, S. & P. A. Barbosa. 2010. Hesitation phenomena: a dynamical perspective. *Cognitive Processing* 11(3): 251–261. https://doi.org/ 10.1007/s10339-009-0348-x

Pindzola, R. H. 1990. Dysfluency characteristics of aged, normal-speaking black and white males. *Journal of Fluency Disorders* 15: 235–243. https://doi.org/10.-1016/0094-730X(90)90004-C

Roggia, A. B. 2012. Eh as a polyfunctional discourse marker in Dominican Spanish. *Journal of Pragmatics* 44: 1783–1798. https://doi.org/10.1016/j.pragma. 2012. 08.010

Searl, J P., R. M. Gabel, & J. S. Fulks, 2002. Speech disfluency in centenarians. *Journal of Communication Disorders* 35: 383–392. https://doi.org/10.1016/ S0021-9924(02)00084-9

Shriberg, E. 2001. To '*errrr*' is human: Ecology and acoustics of speech disfluencies. *Journal of the International Phonetics Association* 31(1): 153–169. https://doi.org/ 10.1017/s0025100301001128

Silber-Varod, V., H. Kreiner, R. Lovett, Y. Levi-Belz, & N. Amir. 2016. Do social anxiety individuals hesitate more? The prosodic profile of hesitation disfluencies in Social Anxiety Disorder individuals. *Proceedings of Speech Prosody 2016 (SP2016)*. Boston. https://doi.org/ 10.21437/SpeechProsody.2016-249.

Smith, V. L. & H. H. Clark, 1993. On the course of answering questions. *Journal of Memory and Language* 32: 25–38. https://doi.org/10.1006/jmla.1993.1002

Trouvain, J., C. Fauth, & B. Möbius. 2016. Breath and non-breath pauses in fluent and disfluent phases of German and French L1 and L2 read speech. *Proceedings of Speech Prosody 2016*, Boston, 31–35. https://doi.org/ 10.21437/speechprosody.2016-7

Urizar, X. & A. G. Samuel. 2014. A corpus-based study of fillers among native Basque speakers and the role of zera. *Language and Speech* 57(3): 338–366. https:// doi.org/10.1177/0023830913506422

Watanabe, M., K. Hirose, Y. Den, & N. Minematsu. 2008. FPs as cues to the complexity of upcoming phrases for native and non-native listeners, *Speech Communication* 50(2): 81–94. https://doi.org/10.1016/j.specom. 2007. 06.002.

Zellner, B. 1994. Pauses and the temporal structure of speech. In: E. Keller (ed.): *Fundamentals of speech synthesis and speech recognition*. John Wiley, Chichester, 41–62.