Adaptation to natural fast speech and time-compressed speech in children

Hélène Guiraud [1]*, Emmanuel Ferragne*[2]*, Nathalie Bedoin*[1]*, Véronique Boulenger*[1]

[1] Laboratoire Dynamique du Langage, CNRS/Université Lyon 2 UMR 5596, Lyon, France
[2] CLILLAC-ARP, EA 3967, Université Paris Diderot, Paris, France

guiraudh@gmail.com, nathalie.bedoin@univ-lyon2.fr,
emmanuel.ferragne@univ-paris-diderot.fr, veronique.boulenger@ish-lyon.cnrs.fr

Abstract

Adaptation to artificially time-compressed speech and natural fast speech has been previously shown in adult listeners, with significant improvement of performance within 5-10 sentences. In the present study, we investigated whether typically developing children also adapt to such variations in speech rate. Eighteen children performed a semantic judgment task on normal speed sentences, natural fast sentences and time-compressed sentences. The three speech rate conditions were presented in separate blocks to examine adaptation over exposure time. Analysis of response times broken down into miniblocks of 5 sentences of the same rate reveals that whereas performance for normal sentences remains stable over time, response times become significantly shorter after listening to the first 5 sentences, both in the natural fast and time-compressed conditions. Therefore, children find it more difficult to understand natural fast and time-compressed sentences as revealed by increased response times, but after listening to 5 sentences, their performance improves and becomes comparable to that for normal sentences. These preliminary results suggest that children adapt to speech rate changes as rapidly as adults and that they adapt to both types of speech distortion (natural fast and time-compressed) in the same way.

Index Terms: speech perception, children, natural fast speech, time compression, adaptation.

1. Introduction

Every listener has experienced the difficulty of understanding a conversation with someone speaking very fast or having a foreign or regional accent and has noticed that, over time, comprehension becomes much easier. This gradual, yet rapid adaptation to speech variations illustrates the remarkable flexibility of our perceptual system to continuously normalize for changing parameters to eventually understand the delivered message. Probably one of the most frequent variations in speech is the rate at which it is articulated. Increasing speech rate elicits more phonetic phenomena such as coarticulation and assimilation than speaking at a normal rate, and it augments the likelihood of segment deletions [1, 2]. Furthermore, speech rate is increased nonlinearly: in English and Dutch for instance, the durations of vowels and unstressed syllables are reduced more than the durations of consonants and stressed syllables respectively [3-5]. These timing changes and extra amount of segmental overlap that characterize natural fast speech may be particularly challenging for the listener, demanding more processing time and resources.

Previous studies have shown that young and older adults are perfectly able to adapt to artificially time-compressed speech, even for compression rates up to 65 % of the original duration of the signal (~1.5 times normal rate) [5-10]. Dupoux and Green [6] asked young adults to listen to 20 sentences time-compressed at 38% or 45% of their original duration and to write down as much of their content as they could recall. Analysis of recall accuracy (with data broken down into 4 sets of 5 sentences) revealed improving performance across the first sets of items, namely within 5-15 sentences. In two subsequent experiments, the authors further showed that adaptation transferred from one talker to another and was not affected by an intervening shift in compression rate. Although there was a local decline in recall accuracy for the first two sentences immediately following the shift in talker or compression rate, performance rapidly recovered over the last three sentences in the set. The adjustment after a change in talker or compression rate was therefore much more rapid than during the first exposure to time-compressed sentences (i.e. performance was not reset to baseline), suggesting the existence of a long-term adjustment reflecting a perceptual learning process that would operate at a rather abstract level of representation. Other studies have confirmed the existence of rapid adaptation to time-compressed speech in adults and have even shown different patterns of performance for different types of time compression [5, 7]. A processing advantage of linearly time-compressed speech was indeed observed over time-compressed speech that followed the temporal pattern of natural fast speech. Hence, making time-compressed speech closer to the more salient prosodic pattern of natural fast speech had a detrimental effect on intelligibility and ease of processing, compared to linear time compression.

Perceptual adaptation to time-compressed speech has often been suggested to reflect an attention-weighing process in which listeners shift their attention toward task-relevant cues and away from task-irrelevant cues [11, 12]. Adaptation is further assumed to occur as listeners recalibrate phonemic boundaries to accommodate more rapid speech rates [9]. When listening to time-compressed sentences, speech tokens are too distant from the phoneme prototypes contained in normal speech and this would negatively affect identification with such distorted speech. Through exposure, learning of new acoustic representations of stored phonemic categories would however occur and these tokens would become closer in perceptual space [13], leading to improved performance.

In most studies on adaptation to speech rate changes, artificially time-compressed speech has been used. Although it has provided important information on adaptation processes at work during speech perception, an investigation of adaptation to natural fast speech was critically needed. Contrary to time-compressed speech which only alters the temporal structure of the signal, increasing speech rate naturally elicits both temporal and spectral changes and increases the likelihood of non-canonical phonetic realizations stemming from coarticulatory phenomena, which may be more problematic for the listener. Natural fast speech is indeed less carefully articulated than normal speech that has been time-compressed afterwards, and even when the compression rate matches the fastest rate which speakers can achieve, it is still almost perfectly intelligible [5]. This may cause natural fast speech to be processed with more difficulty, in spite of its naturalness.

Two studies, as far as we know, have examined adaptation to natural fast speech in adults [7, 10]. In a sentence verification task, Adank and Janse [10] compared performance as a function of exposure time to time-compressed speech and natural fast speech. They further investigated whether there was transfer of learning of one speech type to the other by comparing performance of two groups of subjects. The first group listened to natural fast speech before time-compressed speech, while the second group was presented with the reverse order (both groups first heard normal rate sentences). Analyses of accuracy and response times (RTs; with data broken down into miniblocks) revealed lower accuracy and longer RTs for natural fast than for normal sentences, whereas the difference between time-compressed and normal sentences only showed up on RTs. Performance further improved over the first 2 miniblocks (especially for the first group) in terms of accuracy and to some extent RTs for natural fast sentences. For time-compressed sentences, improvement (over the first 2-3 miniblocks) was observed only on RTs. Finally, the group of subjects who heard time-compressed speech first performed better for natural fast sentences than the other group, suggesting transfer of learning from adaptation to time-compressed speech to naturally-produced fast speech (the reverse was not true). Overall, these findings demonstrate that although the processing of natural fast speech seems more difficult than that of artificially time-compressed speech, adult listeners are able to adapt to this natural increase of speech rate, and adaptation occurs extremely fast and as rapidly as for time-compressed speech.

To our knowledge, no study to date has examined adaptation processes to varying speech rates in children. This is a critical point as children, just like adults, face different talkers (*e.g.* at home, at school, etc.) speaking at different rates and have to constantly normalize for these changes. Examining whether and how children adapt to speech rate variations is therefore of fundamental importance as (i) it will allow better understanding of the cognitive processes underlying speech perception and (ii) it has clear academic consequences, especially for non-native children who may find it difficult to deal with speech rate changes. Previous works with young infants, aged 2-4 months, have demonstrated that they show sensitivity to syllable duration, which reflects speech rate, in discrimination tasks on stimuli that differed on transition duration [14, 15]. This suggests that adjustment to rate-induced variability could already occur at early stages of development.

In the present experiment, we investigated whether typically developing children adapt to (i) linearly time-compressed speech and (ii) natural fast speech in a sentence semantic judgment task. In agreement with previous studies in adults [6, 7, 10], we expected children to show perceptual learning for the two types of distorted speech. Given that perceptual abilities of children are still not fully developed, adaptation could however occur after exposure to a larger number of sentences than in adults. Children could also show more difficulty to process natural fast speech than time-compressed speech, as has been observed in adults.

2. Materials and Methods

2.1. Participants

Eighteen healthy children, aged 8-9 years (mean = 8.23 years old, SD 0.39), participated in the experiment. All were French native speakers and right-handed with no known hearing or language disorders. Children's language abilities were verified with the French reading test 'L'Alouette' [16]. Their parents signed a consent form. The study was approved by the French Ethics Committee and received support from regional services of the National Education of the Academy of Lyon.

2.2. Stimuli

Three hundred sentences (7–9 words) were created following the same syntactic structure: Determiner – Noun 1 – Verb – Determiner – Noun 2 – Preposition – Determiner – Noun 3. Semantic plausibility of Noun 3 (disyllabic target word) within the sentence context was manipulated so that for half of the sentences, the target word was congruent with the context, whereas for the other half it was incongruent. Each target word appeared both in a congruent and an incongruent context (*e.g.* "Sa fille déteste la nourriture de la cantine" and "Le public applaudit le joueur pour sa cantine" (*His daughter hates the food at the canteen* and *The public applauds the player for his canteen*). In all cases, the semantic congruent/incongruent content of the sentence was obvious.

Sentences were recorded by a French native male speaker (32 years-old; 44.1 kHz, mono, 16 bits) in a sound-attenuated booth, using ROCme! software [17]. Each sentence was recorded twice, at normal and fast speech rate. The procedure was the following: the sentence was first presented on a computer screen in front of the speaker. He was instructed to silently read the sentence and to subsequently produce it as a declarative statement at his normal speech rate. Next, the sentences were produced at a fast speech rate using the same procedure. The speaker could produce each sentence several times so that the recorded version was as fluent as possible.

The durations of the 2×300 sentences and the number of produced syllables for each sentence were then calculated. The average speech rate was 6.76 syllables/s (SD 0.57) for the natural normal sentences and 9.15 syllables/s (SD 0.60) for natural fast speech sentences. Thus, the overall fast-to-normal ratio was 0.74 (i.e. speed-up factor of 1.35). Subsequently, the time-compressed sentences were computed by digitally shortening them with a PSOLA algorithm as implemented in Praat. Compression rates were obtained for each sentence: each individual time-compressed sentence was matched in terms of rate to its corresponding natural fast item. For the total of 900 sound files (300×3 variants), an 80 Hz high-pass filter was applied and the amplitude envelope was smoothed sentence-initially and finally. The intensity of the sound files was finally peak normalized.

The 300 sentences were divided into 12 experimental lists of 75 items each using a Latin square design; each list was composed of 3 experimental blocks (25 items, 13 congruent/12 incongruent) corresponding to the 3 speech rate conditions. The blocks were always presented in the same order, namely normal sentences, natural fast sentences and time-compressed sentences (see [10]). Each list contained every sentence only once to avoid repetition effects, and no congruent-incongruent sentence pairs were presented within the same list and thus to each participant. Final target words were matched for word frequency, number of phonemes and number of phonological neighbors between lists using the French lexical databases Lexique [18] and Manulex [19]. Across lists, all target words were presented against the 6 different conditions (3 speech rates × 2 semantic congruency). Within each experimental block, the order of sentences was randomized across participants.

2.3. Procedure

Children were tested in a silent room at school and received oral instructions. They were comfortably seated facing a laptop. Stimuli were presented diotically over headphones at a comfortable sound level. Children were instructed to attentively listen to the stimuli and to perform a semantic judgment task, that is, they had to decide as quickly and accurately as possible whether the sentences made sense or not by pressing one of two keys on the keyboard. Once the children gave their response, the next trial was presented after 1 s. If the children did not respond within 7 s, the trial was recorded as "no response" and the next trial was presented. Participants could listen to each stimulus no more than once. Before the testing phase, they were given 5 practice items (different from the experimental stimuli and produced by the same speaker). A short break was proposed after each block. The total duration of the experiment was 25 minutes. Stimulus presentation, response time and error measurements were performed using E-Prime 2 (Psychology Software Tools, Inc., Pittsburgh, PA).

2.4. Statistical analysis

Response times (RTs; time-interval between the onset of the target word and the button press; in milliseconds) and accuracy (% of correct responses) were measured. Trials for which participants made no response or made mistakes were considered as errors and were not included in the RT analysis. Trials with RTs below or above 2.5 standard deviations from individual means were further excluded from the RT analysis.

In order to examine adaptation over exposure time, we split each experimental block (i.e. each speech rate condition) into 5 miniblocks of 5 sentences. Mean RTs and accuracy were used as dependant variables in separate statistical analyses. Two-way repeated measures analyses of variance (ANOVAs) were conducted, with Speech Rate (normal vs. natural fast vs. time-compressed) and Miniblock (1 to 5) as within-subjects factors and Subject as a random factor. Since we had specific predictions about the time-course of adaptation, which should occur rapidly (within 5-10 sentences, [6]), we also performed the two-way repeated measures ANOVAs on the first 2 miniblocks only for each speech rate condition. In case of significant interactions, paired *t*-tests were used to examine differences between conditions.

3. Results

Data from 4 children were excluded from the analysis as their average RTs were more than 2.5 standard deviations slower than the average across participants.

3.1. Response times analysis

Figure 1 displays RTs averaged over participants for each speech rate condition, broken down into 5 miniblocks of 5 sentences. As illustrated by this figure, RTs seemed generally longer in the natural fast and time-compressed conditions than in the normal rate condition. In addition, RTs in the first miniblock of natural fast and time-compressed sentences were longer than in the 4 subsequent blocks of the same conditions, whereas performance tended to be stable across miniblocks for normal sentences.

The ANOVA (Speech Rate × Miniblocks) did not reveal any significant main effects, nor there was any interaction between the two factors. The RT increase for natural fast (mean = 1737 ms, SD 619) and time-compressed sentences (1759 ms, SD 550) compared to normal sentences (1658 ms, SD 511) seen in Figure 1 was thus not confirmed by statistical analysis. Similarly, the lack of a significant main effect of Miniblock did not confirm the observation from Figure 1. However, when only the first 2 miniblocks were considered in the statistical analysis, a significant main effect of Miniblock was found ($F(1, 12) = 7.953$, $p = .015$) as well as a significant Speech Rate × Miniblock interaction ($F(2, 25) = 3.646$, $p = .040$). For the natural fast and time-compressed sentences, average RTs in the first miniblock of 5 sentences (1890 ms, SD 649 and 1975 ms, SD 631 respectively) were significantly longer than in the second miniblock (1632 ms, SD 459 and 1623 ms, SD 558 respectively). No such difference was observed for normal rate sentences (1643 ms, SD 472 for miniblock 1 vs. 1643 ms, SD 548 for miniblock 2). Paired *t*-tests confirmed that performance significantly improved between the first and the second miniblocks of natural fast sentences ($p = .042$) and time-compressed sentences ($p = .010$). Children became 258 ms faster between the first two miniblocks of natural fast sentences and 351 ms faster between the first two miniblocks of time-compressed sentences.

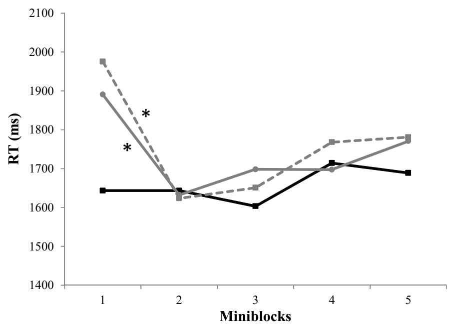

Figure 1: *Average RTs (in ms) per miniblock of five sentences for the normal rate condition (black line), natural fast rate condition (grey line) and time-compressed condition (dotted grey line). (*) indicates a significant difference between conditions.*

3.2. Accuracy analysis

The ANOVA (Speech Rate × Miniblocks) revealed a significant main effect of Speech Rate ($F(2, 198) = 5.62$, $p = .004$). Percentage of correct responses was significantly lower for natural fast sentences (60%, SD 27.2) than for normal (82.8%, SD 25.8) and time-compressed sentences (75.7%, SD 19.5) which did not significantly differ from each other (Figure 2). No significant effect of Miniblock and no interaction between the two factors were observed. This suggests that accuracy did not improve as a function of exposure to sentences in any of the speech rate conditions.

Finally, the ANOVA performed on the first two miniblocks only did not reveal any significant main effects of Speech Rate and Miniblocks nor any interaction between the two.

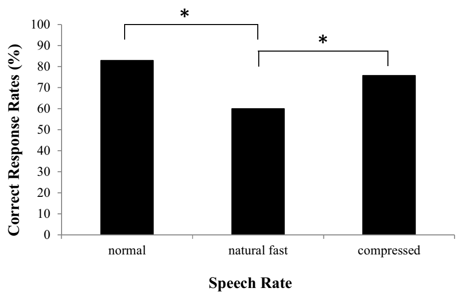

Figure 2: *Mean accuracy (percentage of correct responses) for the normal rate, natural fast and time-compressed conditions. (*) indicates a significant difference between conditions.*

4. Discussion

This experiment investigated adaptation to natural fast speech and linearly time-compressed speech in typically developing 8-9 year-old children in a semantic judgment task. Results reveal that children performed poorer in terms of accuracy for natural fast sentences than for normal and time-compressed sentences, which did not differ from each other. The advantage of artificial time compression (and normal rate) over natural acceleration of speech rate did however not show up on response times (RTs). These results are consistent with the findings of [10] showing lower accuracy and longer RTs in a sentence verification task for natural fast sentences than for normal sentences, and longer RTs (but not lower accuracy, as in our study) in the time-compressed than in the normal rate condition. In a phoneme detection task, Janse [7] also reported increased RTs for natural fast speech compared to time-compressed speech (note that accuracy was not analyzed in [7]). In general, the processing of naturally-produced fast speech therefore seems more demanding than that of linearly time-compressed speech in different tasks and languages, probably owing to its greater spectro-temporal variation and increased segmental overlap [1, 2]. Our preliminary findings suggest that this is also the case in typically developing children.

Our results further demonstrate that children are perfectly able to adapt to (i) time-compressed speech and (ii) natural fast speech, just like adults. This adaptation, as reflected by an improvement of RTs as a function of exposure time, occurred after listening to 5-10 sentences for both speech rate conditions, which is within the range reported in adults [6, 10]. Such an improvement provides evidence that experience with natural fast and time-compressed speech resulted in perceptual adjustment in children. Note that different languages or tasks were used in the different studies; a direct comparison of adaptation to speech rate variations in French children and adults with exactly the same stimuli and task is currently in progress. Finally, our data do not seem to be accounted for by general improvement of performance over the course of the experiment due to practice with the task. If so, we should indeed have observed the same degree of improvement for sentences produced at a normal rate, which was not the case.

It is noteworthy that adaptation occurred similarly for the two types of accelerated speech: performance improved after listening to ~5 sentences, independently of the way speech rate was speeded up, that is, the time to approach a plateau was comparable for the natural fast and time-compressed conditions. Nevertheless, the fact that children's accuracy was significantly lower for natural fast than for time-compressed sentences may reflect qualitative differences in the processes underlying perception of these two types of faster speech. We here propose a tentative explanation that may account for our findings. To deal with the greater spectro-temporal variation and non-canonical phonetic realizations of natural fast speech, the listener may be required to shift from a set of acoustic-phonetic rules towards another set of appropriate parsing rules in order to extract the most relevant cues for phonological identification [9, 12]. Natural fast speech may indeed disturb the typical pattern of perceptual relevance of acoustic cues to phoneme identification and therefore be particularly challenging for the listener. This resource-demanding process may recruit additional brain regions, possibly located in the premotor cortex [20, 21], compared to normal speech (see also [22] for fast speech perception in blind listeners). In children, the set of specific rules that allows coping with fast speech rates may still be imprecise, which could lead to lower accuracy in fast speech perception tasks. Contrary to natural fast speech, linear time compression of normal speech only alters the temporal structure of segments. Accordingly, shorter temporal cues have to be processed than in normal speech, which may specifically reinforce the recruitment of the left hemisphere [23, 24], but may not require to shift to an entirely different set of appropriate rules. This might explain the comparable accuracy in the time-compressed and the normal rate conditions in our experiment.

It is acknowledged that speech development continues even after the establishment of intelligible speech production. Studies have actually demonstrated a gradual increase in speech rate in children at different ages (from 9 to 11) and slower speech rate in children than in adults [25, 26]. Given the close relationship between the speech production and perception systems [27] and the change in the perceptual relevance of dynamic and static cues to consonant identification over the course of development [28], it is conceivable that adaptation to natural fast and time-compressed speech could occur differently in children at different ages. Future studies are required to provide such empirical evidence and to directly compare their performance to that of adults.

5. Conclusions

Our preliminary findings provide, to our knowledge, the first demonstration of adaptation to natural fast speech and time-compressed speech in healthy children. Although future studies are required to strengthen this assumption, this highlights the remarkable flexibility and adaptability of the human speech comprehension system, early in the course of development.

6. Acknowledgements

We thank Damien Gouy for recording the stimuli. We also thank the children and their parents as well as the school Cavenne in Lyon. This study was carried out with financial support from the Agence Nationale de la Recherche (ODYSSEE project; PI: V.Boulenger).

7. References

[1] Byrd, D. and Tan, C. C., "Saying consonant clusters quickly", Journal of Phonetics, 24(2):263-282, 1996.

[2] Koreman, J., "Perceived speech rate: the effects of articulation rate and speaking style in spontaneous speech", J Acoust Soc Am, 119(1):582-596, 2006.

[3] Max, L. and Caruso, A. J., "Acoustic measures of temporal intervals across speaking rates: variability of syllable- and phrase-level relative timing", J Speech Lang Hear Res, 40(5):1097-1110, 1997.

[4] Peterson, G. E. and Lehiste, I., "Duration of syllable nuclei in English.", Journal of the Acoustical Society of America, 32(6):693-703, 1960.

[5] Janse, E., Nooteboom, S., and Quené, H., "Word-level intelligibility of time-compressed speech: prosodic and segmental factors", Speech Communication, 41:287-301, 2003.

[6] Dupoux, E. and Green, K., "Perceptual adjustment to highly compressed speech: effects of talker and rate changes", J Exp Psychol Hum Percept Perform, 23(3):914-927, 1997.

[7] Janse, E., "Word perception in fast speech: artificially time-compressed vs. naturally produced fast speech", Speech Communication, 42:155-173, 2004.

[8] Peelle, J. E. and Wingfield, A., "Dissociations in perceptual learning revealed by adult age differences in adaptation to time-compressed speech", J Exp Psychol Hum Percept Perform, 31(6):1315-1330, 2005.

[9] Golomb, J. D., Peelle, J. E., and Wingfield, A., "Effects of stimulus variability and adult aging on adaptation to time-compressed speech", J Acoust Soc Am, 121(3):1701-1708, 2007.

[10] Adank, P. and Janse, E., "Perceptual learning of time-compressed and natural fast speech", J Acoust Soc Am, 126(5):2649-2659, 2009.

[11] Goldstone, R. L., "Perceptual learning", Annu Rev Psychol, 49:585-612, 1998.

[12] Francis, A. L., Baldwin, K., and Nusbaum, H. C., "Effects of training on attention to acoustic cues", Percept Psychophys, 62(8):1668-1680, 2000.

[13] Iverson, P. and Kuhl, P. K., "Mapping the perceptual magnet effect for speech using signal detection theory and multidimensional scaling", J Acoust Soc Am, 97:553-562, 1995.

[14] Eimas, P. D. and Miller, J. L., "Contextual effects in infant speech perception", Science, 209(4461):1140-1141, 1980.

[15] Miller, J. L. and Eimas, P. D., "Studies on the categorization of speech by infants", Cognition, 13(2):135-165, 1983.

[16] Lefavrais, P., "Test de l'Alouette", Éditions du Centre de Psychologie Appliquée, Paris), 1967.

[17] Ferragne, E., Flavier, S., and Fressard, C., "ROCme! Recording of Oral Corpora Made Easy", 2.0, 2012, available from http://www.ddl.ish-lyon.cnrs.fr/.

[18] New, B., Pallier, C., and Ferrand, L., "Lexique 2: A New French Lexical Database", Behav Res Methods Instrum Comput, 36(3):516-524, 2005.

[19] Lété, B., Sprenger-Charolles, L., and Colé, P., "MANULEX : A grade-level lexical database from French elementary-school readers", Behavior Research Methods, Instruments, & Computers, 36:156-166, 2004.

[20] Peelle, J. E., McMillan, C., Moore, P., Grossman, M., and Wingfield, A., "Dissociable patterns of brain activity during comprehension of rapid and syntactically complex speech: evidence from fMRI", Brain Lang, 91(3):315-325, 2004.

[21] Adank, P. and Devlin, J. T., "On-line plasticity in spoken sentence comprehension: Adapting to time-compressed speech", Neuroimage, 49(1):1124-1132, 2010.

[22] Hertrich, I., Dietrich, S., and Ackermann, H., "Tracking the speech signal--time-locked MEG signals during perception of ultra-fast and moderately fast speech in blind and in sighted listeners", Brain Lang, 124(1):9-21, 2013.

[23] Poeppel, D., "The analysis of speech in different temporal integration windows: cerebral lateralization as asymmetric sampling in time", Speech Communication, 41:245-255, 2003.

[24] Liegeois-Chauvel, C., de Graaf, J. B., Laguitton, V., and Chauvel, P., "Specialization of left auditory cortex for speech perception in man depends on temporal coding", Cereb Cortex, 9(5):484-496, 1999.

[25] Walker, J. F., Archibald, L. M., Cherniak, S. R., and Fish, V. G., "Articulation rate in 3- and 5-year-old children", J Speech Hear Res, 35(1):4-13, 1992.

[26] Smith, B. L., Sugarman, M. D., and Long, S. H., "Experimental manipulation of speaking rate for studying temporal variability in children's speech", J Acoust Soc Am, 74(3):744-749, 1983.

[27] Pickering, M. J. and Garrod, S., "Do people use language production to make predictions during comprehension?", Trends Cogn Sci, 11(3):105-110, 2007.

[28] Ohde, R. N., Haley, K. L., Vorperian, H. K., and McMahon, C. W., "A developmental study of the perception of onset spectra for stop consonants in different vowel environments", J Acoust Soc Am, 97(6):3800-3812, 1995.