Available online at www.sciencedirect.com

ScienceDirect

Speech Communication 75 (2015) 84–96

SPEECH COMMUNICATION

www.elsevier.com/locate/specom

# Too fast to be true? Exploring time compression in simultaneous interpreting

Rafael Barranco-Droege *

*Universidad de Granada, Dep. de Traducción e Interpretación, 18071 Granada, Spain*

Received 4 April 2015; received in revised form 11 September 2015; accepted 17 September 2015
Available online 9 October 2015

**Abstract**

When speaking fast, we tend to reduce sentence-ending pauses, potentially impinging on their function as cues for the listener. Earlier research indicates that fast natural speech is harder to process than time-compressed speech, in which all pauses are reduced proportionally. To check whether this advantage for listening withstands the high cognitive demands of simultaneous interpreting, we set up an experiment using two videotaped versions of a content-rich speech. One was delivered by the speaker at 166 words/min and the other one was an originally slower presentation made similarly fast using a 20% compression rate. We asked eight professional practitioners to interpret both versions and to estimate their interpreting difficulty in terms of several dimensions. When the two versions were presented separately, the subjects perceived the time-compressed recording as harder to interpret with regard to *linguistic expression*. When comparing them, they perceived the time-compressed version either as similarly or as more difficult to interpret than the fast natural version. These results cast doubts on the conjectured advantage of time compression for the interpretation of content-rich speeches.


*Keywords:* Simultaneous interpreting; Speech perception; Time-compressed speech; Fast natural speech

## 1. Introduction

This study was carried out within a broader research project. Its aim was to test two methods for the treatment of speaking speed as an independent variable.

The most straightforward method to achieve two versions of a speech at different speaking speeds consists in letting a speaker deliver the same script twice and recording both performances. A trained speaker, provided with adequate supervision, may be able to reproduce prosodic patterns with great accuracy. However, higher speaking speeds seem to lead inextricably to the rise of certain phenomena. On one hand, time constraints may lead to poor articulation and even to phoneme deletion (Koreman, 2006). On the other hand, in fast speech, long pauses are shortened disproportionately, in particular those located at syntactic boundaries (*juncture pauses*). Speakers try to compensate for this by resorting to variations in articulation rate and intonation (Strangert, 1991; Janse et al., 2003) and by avoiding word-internal interruptions (Tydgat et al., 2011). This complex interplay of factors is made apparent by the fact that the pauses identified by listeners do not necessarily match actual pauses (Megyesi and Gustafson-Čapková, 2002). In sum, natural-speed manipulation is accompanied by confounding effects associated with local variations in pause patterns, rhythm and intonation. Therefore, this method does not seem appropriate.

These shortcomings could be overcome by choosing an alternative manipulation method that scales temporal patterns evenly (durations of sound and silence, speed variations) and preserves the prominence patterns of the original recording (intonation, stress). These requirements

* ORCID: 0000-0002-2699-8337.
*E-mail address:* rafael@rafaelbarranco.com

http://dx.doi.org/10.1016/j.specom.2015.09.009
0167-6393/

are fulfilled by *linear time manipulation*, which consists either in *time expansion* (slowing down) or in *time compression* (speeding up).[1] Today there are various software tools that make this possible without transposing pitch (no "chipmunk voice"). Indeed, Janse (2004) showed that fast natural speech is more difficult to process by listeners than speech that has been linearly time-compressed to achieve the same speech rate. She attributed this result to the reduction of juncture pauses and to the less careful articulation in fast natural speech.

Moderate time expansion can improve intelligibility in noisy environments (Wenndt et al., 1996). Moreover, listeners can adjust to time-compressed speech, although this ability does not seem to carry over between languages (Pallier et al., 1998). The time required for adjustment increases with compression rate (Dupoux and Green, 1997), and listening to time-compressed speech may induce a preference for higher listening rates (Gade and Mills, 1989). Quite surprisingly, in a study with older listeners, it was found that the acoustic recognition of time-compressed speech did not predict recognition of fast natural speech (Gordon-Salant et al., 2014).

In addition to acoustical perception and comprehension, simultaneous interpreting involves other cognitive tasks. Gile's Effort Model (e.g. 2009) describes the mental processes of professionals in this interpreting mode in terms of four cognitive *efforts*: *listening* and analysis, *memory management*, *production* and *coordination*. The weighting of each of these efforts is determined by the source speech, the communicational situation and the interpreter's strategies, and the sum of these efforts is limited by the total cognitive capacity available to the interpreter at any given moment.

In this study, we wanted to find out whether time-compressed ('artificial') speech is easier to interpret than 'natural' speech delivered at a similarly fast rate. More specifically, we wanted to explore the limits of time compression. This implied choosing a high compression rate that fell short of being immediately noticeable to listeners and a fast rate that still felt comfortable to the speaker.

For this purpose, we let a group of professional interpreters evaluate the difficulty of two versions of a source speech: a fast natural version and a time-compressed version with a similar speech rate. Their feedback would hopefully provide answers to the following research questions:

- RQ1. Which one is perceived as more difficult to interpret, the fast natural version or the time-compressed version?
- RQ2. Which one is perceived as more difficult to interpret when compared with the other one, the fast natural version or the time-compressed version?

The second question was introduced presuming that the scrutiny of order effects may be useful to understand the reasons behind first impressions. Order effects have been found in the perceptions of interpreting users (García Becerra, 2012).

## 2. Material and methods

This section describes the subjects, the interview format, the experimental procedure and the methods of analysis. It goes on to present the audiovisual materials, carefully considering a number of confounding variables.

### 2.1. Subjects

Eight subjects, seven women and one man, were recruited from a group of interpreters taking part in a course for interpreter trainers in Germany. At that event, roughly half of them acted as participants and the rest worked as teachers. We invited them individually to participate in an experiment aimed at "eliciting their perceptions as professional interpreters". We informed them that they would be asked to interpret two recordings and to fill out a questionnaire on each. To induce them to focus on the listening task, we told them that their utterances would not be recorded. During the experiment, they would be seen, but not heard from outside the booth. All of the persons contacted agreed to participate.

The sociodemographic data compiled during the experiment show that the subjects' professional experience was diverse: two had less than six years of experience as conference interpreters, two between six and 10, one between 11 and 15 and three more than 20. The same holds for their teaching experience: two subjects had less than six years of experience as conference-interpreting teachers, two between six and 10, two between 11 and 15 and two more than 20. One of them is a native speaker of English, and the rest are German natives with English as a B (active) or C (passive) language. Half of them were between 30 and 44 years old, and the rest were between 45 and 60 years old.

### 2.2. Interview format

After asking each of the participants several sociodemographic questions, we gave them a 326-word handout written in English, containing general information about the topic of the press conference from which the speech was taken. Then, we invited the subjects to interpret into German the two versions of a speech in English described

[1] Time compression and time expansion might not be equally appropriate as manipulation methods. Our experience suggests that time expansion tends to impinge on timbre and to create artifacts (e.g. within filled pauses), and it may lead to a negative judgment of rhetorical quality (Barranco-Droege, 2015). This may be because it involves a reduction in sampling rate. For its part, time compression can make short phonemes imperceptible, lead to unconvincingly fast articulation (e.g. of consonant clusters) or produce discontinuities (such as click sounds and abrupt stops). However, we have also noticed that it allows higher manipulation rates than time expansion before distortions become noticeable.