S. Diefenbach, N. Henze & M. Pielot (Hrsg.): Mensch und Computer 2015 Tagungsband, Stuttgart: Oldenbourg Wissenschaftsverlag, 2015, S. 413-416.

A Voice Driven Type Design Demo

Matthias Wölfel[1], Angelo Stitz[2], Tim Schlippe[3]

School of Digital Media, Furtwangen University, Germany[1]
School of Design, Pforzheim University, Germany[2]
Research Karlsruhe, Germany[3]

Abstract

With *voice driven type design* (VDTD), we introduce a novel concept to present written information in the digital age. While the shape of a single typographic character has been treated as an unchangeable property until today, we present an innovative method to adjust the shape of each single character according to particular acoustic features in the spoken reference. Thereby, we allow to keep some individuality and to gain additional value in written text, which offers different applications – providing meta-information in subtitles and chats, supporting deaf and hearing impaired people, illustrating intonation and accentuation in books for language learners – up to artistic expression. In this paper we describe the demo system as demonstrated at the conference Mensch und Computer 2015.

1 Background

With the introduction of the movable-type printing system in Europe by Johannes Gutenberg around 1450s, movable types could demonstrate their superiority. After their invention in the 1860s, typewriters became a convenient tool for practically all written communication and quickly replaced handwriting except for personal correspondence. While industrialization necessitated standardization of type in that replication process, digitization offers a liberation of these stringent formats: Fonts have been developed in all kind of flavors. But keys, since the invention of the typewriter, stayed the preliminary input modality. This has also not been changed, in the late 1960th, by the invention of the word processor. Speech as an alternative form of input modality, in contrast to a keyboard, contains more information. This is complementary to text and reflects individuality and emotion. However, it is simply thrown away and not used to influence the type design or to represent in other graphical form. We got so used to this generic transfer of information that nobody challenges this way of visualizing information. In order to keep the additional information such as emotion, prosody and personalization present in verbal communication also in text-based communication, we introduce the novel concept which we have dubbed *voice driven type design*.

2 Voice Driven Type Design

We propose to vary the shape of a single character according to particular acoustic features in the spoken reference. Our motivation of a grapheme level adaptation of the transcription is to better represent the characteristics of the spoken utterance and to keep individuality in written text. This strategy allows additional meta-information in subtitles and chats, supporting hearing impaired and deaf people, illustrating intonation and accentuation in books for language learners as well as artistic expression – which would be more limited or even not possible with previous approaches.

To express a bandwidth of emotional states it necessitates that the origin form, from which the dynamic character shape is formed of, must constitute a simple, generic, rational and reduced basic form. Therefore, we adopted one of the most satisfactory, modernist sans serif typefaces of the 20th century "Futura" by Paul Renner (1927).

2.1 Algorithm

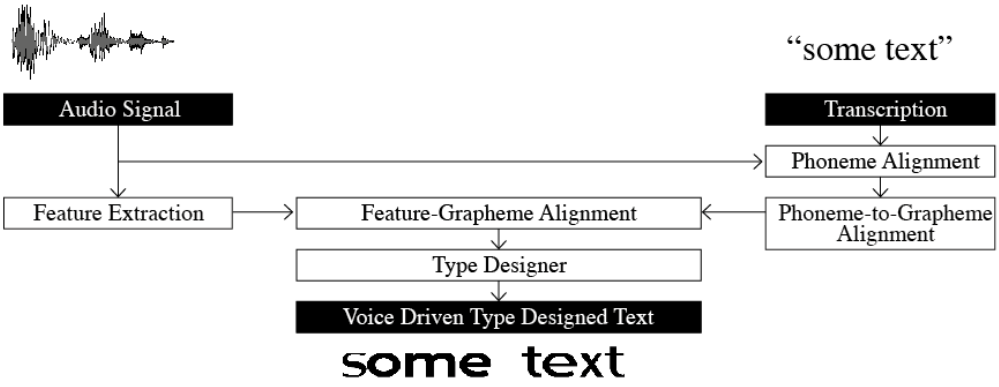

Figure 1: From speech signal to voice driven type designed text.

Given a spoken utterance and its transcription VDTD can be generated by the following steps, which are illustrated in Figure 1:

1. Phoneme alignment and acoustic feature extraction
   - Generate phoneme transcription and determine beginning and end of each phoneme
   - Determine *loudness* and *pitch* (step size 10 ms) given the acoustic signal
2. Phoneme-to-grapheme alignment
   - Align each phoneme to one or more graphemes
3. Features-grapheme alignment
   - Determine speed parameter of the grapheme by using the beginning and end times of the corresponding phoneme

- Determine *loudness* and *pitch* parameters of the grapheme by averaging *loudness* and *pitch* according to the beginning and end times of the corresponding phoneme or phonemes

4. Type design:
   - Generate the shape of each character according to the corresponding normalized (mean and variance) features *loudness*, *pitch* and *speed*

2.2 Mapping Voice Characteristics to Character Shape

Adding values present in acoustic signals into a visual representation makes only sense if it can be interpreted to 'extract' the original meaning. Therefore, we mapped the voice to the character shape based on common principles in formatting speech and typography as shown in Figure 2:

- *Loudness:* Producing loudness in speech amplifies the signal and is usually used to have the attention of a listener. To have the attention of a reader, bolder text – produced with more stroke weight – is commonly used to recognize important keywords in written text.
- *Speed*: A reader usually jumps from a part of a word to a next part of a word. Increasing the character width extends the time of this scanning process of the eyes. Therefore, we map the speed of the utterance to the character width.
- *Pitch:* Movements of pitch levels reflect the emotional expressions of joy, anxiety, or fear in utterances, while medium pitch levels account for more neutral attitudes. High stroke contrast invites a reader for a lower flow of reading and for looking consciously at the character itself.

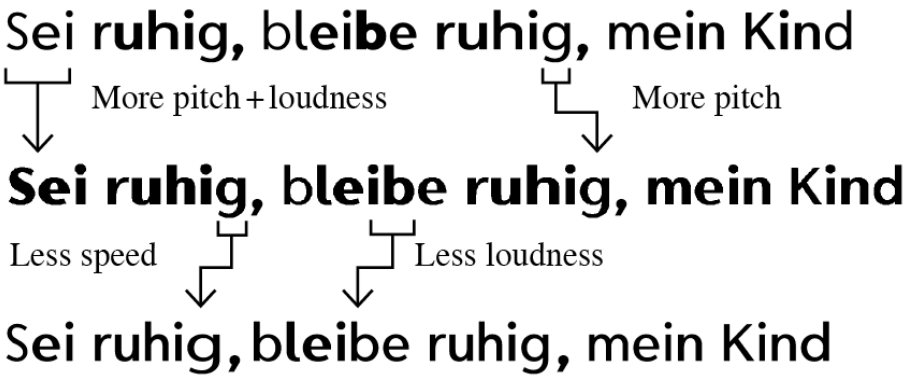

Figure 2: Individual speech characteristics are also visible in spoken text due to VDTD.

3 Demo Setup

The demo lets the user experience VDTD by speaking an utterance into a microphone. The text is generated by VDTD and presented to the speaker.

Figure 3 illustrates our demo setup. Emotional text is displayed on our monitor (2). This text is read by a person and recorded with a microphone (1). The speech signal is sent to our laptop (3), where it is converted to VDTD based on our algorithm. Finally, the redesigned text is displayed on the monitor (2). The person and further spectators are able to see the characteristics of the spoken text on the monitor.

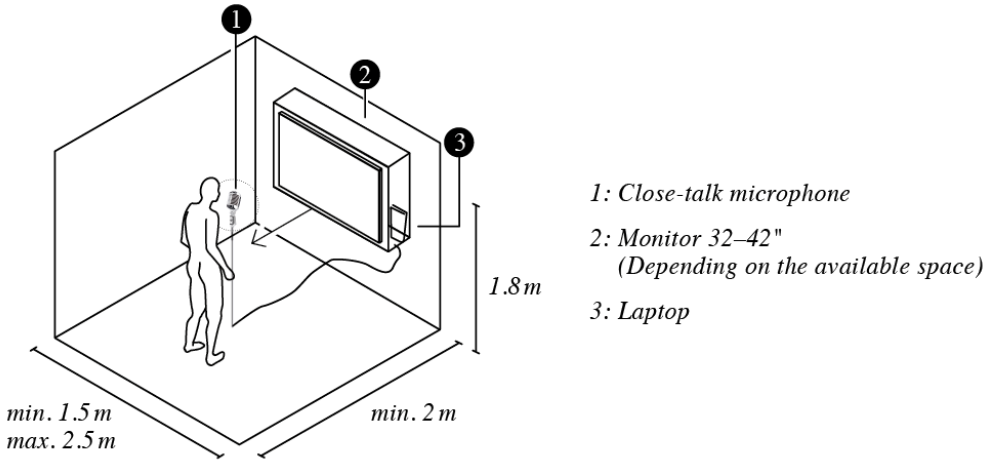

Figure 3: Demo Setup

We decided to use a close talk microphone to avoid the recording of background noise. The size of the monitor can be selected depending on the available space. A larger monitor enables to reach more spectators. A projector and a projector screen are also possible. A laptop is sufficient for our demo because our approach does not require extraordinary computation. Our current system is implemented on a Linux operation system.

The text to read can be selected based on the topic of the conference or exhibition. However, we suggest using content that can be read with varying *loudness*, *speed* and *pitch*. For example poems such as Wolfgang Goethe's famous German poem "Erlkönig" (Erl-King) provide interesting visualizations.