Prosodic Writing with 2D- and 3D-fonts: An approach to integrate pronunciation in writing systems

Markus Rude

Abstract:
Acquisition of intonation and rhythm of a foreign language – e.g. that of English for Japanese language learners – is a challenge, since text-based teaching materials hardly contain prosodic cues. Punctuation gives only insufficient and sometimes misleading indications. The research question is: what kind of script is optimum for a language learner? More specifically: what writing system enables language learners to extract easily maximum linguistic – including prosodic – data from a learner text, while preparing them to read standard texts? A solution is suggested: Prosodic writing (PW) is a writing system that encodes prosodic data in the 3D geometry of letter strings (e.g. height expresses pitch) that otherwise follow standard orthography. The key argument: evolution optimized human cognition to quickly perceive and analyze 3D geometrical data from visual scenes. After introducing the concept, some experimental results are presented, for example the impact on learners' sentence final intonation. Finally, the employment of 3D fonts for PW (3D letters rather than just 2D letter strings floating in 3D space) is suggested: they employ rich potential for encoding additional data into learner texts while being readable with only modest additional cognitive load.
Keywords: writing systems, allograph, typography, readability, prosodic writing, 3D font, phonology, phonetics, parallax, stereogram, 3D vision, perspective.

For Kirsten and Yoshi.

1. Introduction

Foreign language learners find plenty of text books for studying foreign languages. However, concerning pronunciation and prosody, effective teaching materials in visual form are seldom included in general textbooks, therefore dedicated pronunciation text books (e.g. Gilbert 2001) need to be used. Prosodic writing (PW) is an attempt to visualize some of the various dimensions of pronunciation, in particular the 3 prosodic dimensions, *pitch*, *loudness*, and *timing*. PW strives to visualize auditive features in an intuitive way, while staying as close as possible to standard orthography. The familiar forms of letters, words and sentences are mapped into and smoothened in 3-dimensional (3D) geometric space, the dimensions of which correspond to the 3 prosodic dimensions. Such learner texts might look similar to standard texts, and henceforth they could be included in general text books; yet they would include prosodic information that is usually invisible, and thus help non-native learners to develop a sense for rhythm and intonation that is often lacking for too long in the language acquisition process (Fig. 1). Native speaker children already have such a sense when they start to learn reading or writing, therefore general orthography does not need to encode prosodic cues.

The main hypothesis: visual perception of the 3D-form of words and sentences helps learners to grasp the 3D-prosodic form of utterances, namely *pitch phenomena*, *stresses*, and their *temporal patterns* (resulting in intonation contour and rhythm), to use it for reading comprehension (e.g. stress often indicates key information), for reading out, memorizing and reproducing more properly, and for gradually building up their own prosodic systems.

The second hypothesis: viewer-readers are able to synchronously process symbolic data (text as a sequence of standard characters coding linguistic data) and geometric data (geometry of graphemes coding pronunciation data); this simultaneous processing is comparable to interpreting a red stop sign that carries both the linguistic message ("STOP") as well as the geometric-iconic message (red hexagon or triangle); thus, viewer-readers will be able to perceive and interpret relevant information more efficiently than from writing systems with only symbolic prosodic cues.

The main argument for both hypotheses is the evolutionary fact that human beings have been physiologically optimized during thousands of years to instantaneously recognize, classify and interpret 3D objects, including their position, orientation and motion in space.

In this paper, the theoretical concept of PW is explained (section 2) and results obtained so far are reported (section 3). Questionnaires showed the acceptance by students, experiments showed or indicated the effectiveness of the method for several prosodic features. One drawback of the practical application in class is the complexity of producing samples of prosodic writing: Since specific software for the generation of PW is still lacking, the up-to-date applications have only occasionally been produced by computer, mainly they have been generated manually. Whereas the currently applied concept at Nagoya University (section 4) can most easily be imagined as a string of letters (e.g. written on a flexible paper stripe) curving gently in a 3D space, this paper argues that 3D letters in 3D space are still better suited

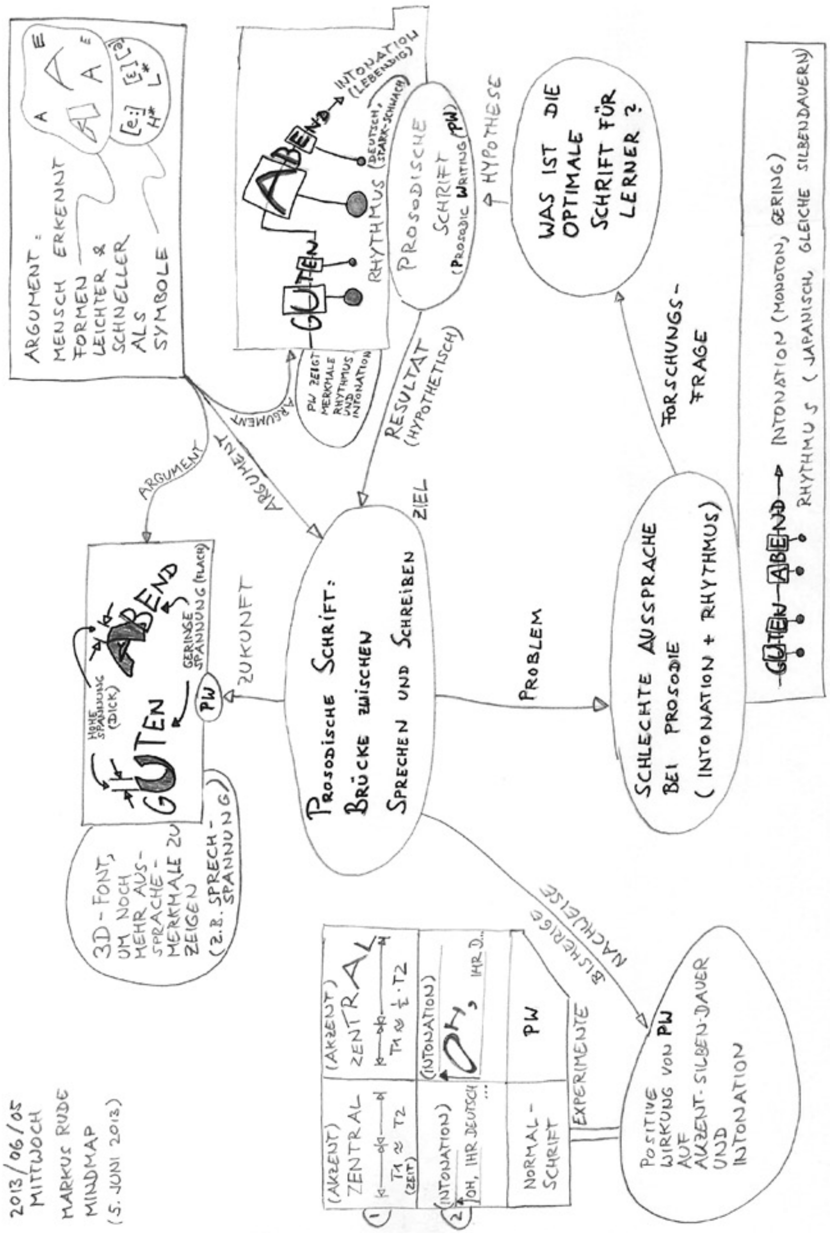

Figure 1: Mindmap showing briefly the contents of this paper. (In German) Problem: Bad pronunciation (Intonation and rhythm). Research question: Optimum script for language learners? Hypothetical result & goal: Prosodic Writing (PW) as a bridge between speaking and writing. Past results: Impact on stress, duration & intonation. Future: 3D-Fonts for PW. Main argument for PW: Human perception is optimized for quickly recognizing spatial forms.

to human perception and cognition (section 5). The potential of coding additional linguistic or para-linguistic data in the text is finally discussed (section 6).

2. Concept of prosodic writing

The optimum writing system for language learners will probably never be found: Is the IPA (International Phonetic Alphabet) a writing system that could be used for language learners? Or should learners immediately read standard texts?

On the one hand, there are still some advocates of using phonetic transcriptions in language classes. In fact, the founders of the IPA (2nd meaning of the same acronym: the International Phonetics Association) wanted to help Parisian school children to develop a good pronunciation of foreign languages through a writing system that captures the pronunciation of foreign sounds better than the Roman alphabet – they developed the IPA. For non-European language learners, however, it is a considerable difficulty to acquire this set of new symbols or new meanings for some symbols they might already know. But even for European language learners familiar with Roman characters on which the IPA is largely based, IPA transcriptions are hard to read and their usage limited to pronunciation practice, since they do also need access to the systematic aspects of morphology, lexicon and syntax; these aspects, however, are much more transparent in texts in standard orthography with its variety of systematic aspects in phoneme-grapheme mappings: e.g. the *principle of homonymy* (different spelling to reflect different meaning despite identical pronunciation, as with "I" and "eye"), the *morphological principle* (identical spelling for etymologically related word(part)s despite different pronunciation (like "crime" and "criminal"), *principles of eugraphy* (obeying certain aesthetical principles, like the avoidance of "ii"), and finally, the *phonetic principle* (faithful transcription of all sounds in an individual utterance, like German "Brot" [bro:t^h] (bread), including phenomena like aspiration) *and the phonological principle* (ideally a one-to-one relationship between phonemes and graphemes, like German "Brot" [bro:t] (bread)). Neglecting the former three, IPA transcriptions only obey a compromise of the latter two, called narrow or broad, depending on whether the phonetic or phonological principle is dominant (see Altmann & Ziegenhein 2010, p. 125 ff for these principles).

On the other hand, however, the advocates of the commonplace orthography – as for example in Fig. 2.1 – neglect the dynamics of form issues in pronunciation, like rhythm or intonational contours, which are crucial for fluent production and for engraining utterances into our memory.

Can we satisfy both points of view – the phonetics stance and the orthographic stance – with one single, easily readable writing system? Possibly, a writing system coined prosodic writing can achieve this goal.

Recently, prosody (pronunciation phenomena mainly beyond segments, therefore also called *suprasegmentals*) has been recognized as being equally important as or even more important than individual sounds (or *segments*) for successful communication. Therefore, PW focuses on prosody and its visualization (section 2 and 3). Segmental aspects like tension will be covered in sections 4 and 5.

Figure 2.1: Hand- vs. typewriting in usual orthography (no prosodic cues)
"Good / evening / Elke (German female first name) / Eber ("boar", here: family name) Standard text does not yield sufficient indications of prosody. Omission of punctuation in the typewritten text and capitalization in both is for the sake of simplicity. Clarity of the concept is given more emphasis than typographical detail.

We could therefore start with an "ideal" definition of prosodic writing:
Prosodic writing (PW) is a writing system that encodes prosody of language faithfully into a string of graphemes through applying a defined distortion to the graphemes and words in order to produce a "spatial code". The graphemes of PW can be elements of any alphabet (in this paper the Roman alphabet; however, it could also be the IPA or Japanese Katakana, a syllable alphabet often used for foreign words).
Using the 1st dimension (x-axis, the horizontal) as time axis, the 2nd dimension (y-axis, the vertical) as pitch axis, and the 3rd dimension (z-axis, depth) as loudness axis, and stretching and gently reshaping the letter string into this 3D space yields both, correct orthography plus the three types of data that mainly constitute prosody. The written syllable's graphemes are spatially positioned (x,

y, z) according to the time-pitch-loudness values of the corresponding spoken syllable's phonemes. By perspective mapping (depth, the z-value, is expressed through size-variation), this meander-like 3D-structure can be represented in a 2D-plane (x, y) for printing or displaying.

In detail:

1st step: For standard writing, each letter is of a defined **width** (also called set-width, Haralambous 2007, 11) depending on its form and neighoring letters through typographical rules; for PW, however, the horizontal placement and width is defined through temporal appearance of the corresponding sound or syllable within the utterance (Fig. 2.2).

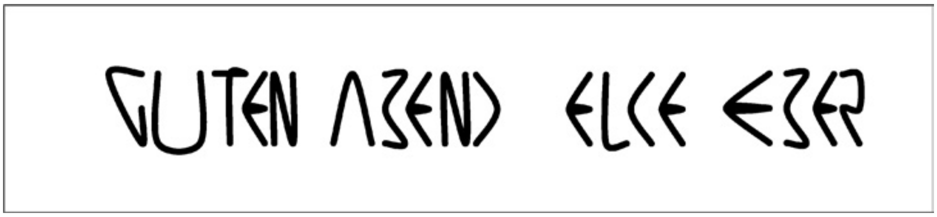

Figure 2.2: Using the x-dimension (the horizontal) as the time axis: Long vowels are stretched.
Long vowels (Guten, Eber) become horizontally stretched. Pauses become more visible through additional horizontal interspace. Note: (1) smaller spaces do not indicate a pause (principle: word separation for readablity). (2) though not being longer in duration, the "G" and "t" around the long "u" in "Guten" become longer in spatial extension (principle: preserving the word form).

The figure already shows a departure from the ideal definition of PW: Word-to-word interspace usually does not have a phonetic correlate. Most adjacent words are linked without audible pauses when spoken. The optimum definition of PW is thus a pragmatic variation of the ideal definition of PW, a compromise between the visualization of prosodic reality, lexical, morphological and syntactical transparency, and overall readability. Similar to IPA transcriptions, we will speak of a broad form (prosodic-phonological, contrastive, idealizing) in contrast to the narrow form (prosodic-phonetical, faithful in its prosodic mapping) from above; for pragmatic reasons, a broad transcription is chosen for this paper and for current PW applications.

In a second step, comparable to musical notation, the vertical displacement of the letters is done according to pitch (Fig. 2.3).

Figure 2.3: Adding the y-dimension (the vertical) for showing pitch
High pitch accent graphemes are positioned higher ("A" in "Abend"), low pitch accent graphemes lower ("E" in "Eber") than the other graphemes. Note: Though "Guten" is not completely flat in an actual utterance, its intonation is phonologically not significant, thus, it is positioned on the horizontal central line.

Again, a suitable learner text should be a compromise between faithful visualization of intonation and readability (e.g. limitation of the inclination angle of letters).

In a third step, loudness is being coded through depth, e.g. on a 3D display. Since paper or an ordinary computer screen nowadays is 2D, real depth cannot be shown: "size" is being used as pseudo-depth: prominent syllables like "Gu" in "Guten", "A" in "Abend" and "E" in "Eber" will be shown enlarged, as in perspective view (objects close to an observer appear bigger) (Fig. 2.4).

Figure 2.4: Adding the z-dimension (pseudo-depth) for showing loudness. (See Fig. 4.2 for perceivable depth) Paper has no depth (z-axis), thus, perspective view is employed: high z-values mean closeness to the observer, the graphemes U, A and E(ber) are bigger. Note: (1) Augmentation here only for accents. (2) As for the graphical correlate of duration (the x-extension of graphemes or syllabic grapheme groups), also the graphical correlate of loudness (z-extension) can involve neighboring graphs to be augmented for the sake of readability (preserving the word form). For example, the graphemes G and T surrounding the U are also slightly enlarged.

In Table 1, the three visual correlates of prosodic features accumulating from Fig.

2.2 to 2.4 are summarized.

Table 1: Summary of prosodic dimensions (left); and the corresponding visual correlates or dimensions (right).

time (t)	x, horizontal
pitch (P)	y, vertical
loudness (L)	z, depth (size as "pseudo-depth")
Summary: Prosodic dimension: [1] [2] [3] [t] [P] [L]	Visual dimension: [1] [2] [3] [x] [y] [z]

Summary:

Prosodic writing is thus a code of spoken language on two levels: On the 1st level, a grapheme is a symbolic code of a phoneme (according to the ideal phonological principle) or a grapheme group is a code of a phoneme group (for non-ideal grapheme-phoneme correspondences) of spoken text, as in ordinary orthographic text; on the 2nd level, each of the written syllable's graphemes carries a geometric code – the above mentioned prosodic parameters – characterizing the corresponding spoken syllable's phonemes; the encoded data components are the three coordinates x, y and z of its position, their meaning is time (of phoneme production), pitch and loudness, respectively, but approximately rather than precisely in a way that is optimum for a learner.

3. Results achieved so far and usage at Nagoya University

PW has been developed since the end of the 1990s by the author during English and German language classes at Japanese universities. From various experiments, one result will be briefly reported in this section, two others will be mentioned.

3.1 Some experimental results

In one experiment at Tsukuba university, students who prepared for an oral test by using PW (learning a dialog by heart) were able to replicate the nuclear contour (intonation contour from the nucleus, the main stress in an utterance, to the end of the utterance) of three target sentences better than a control group that received the text in normal form plus the audio recordings for preparation. (See Fig. 3.1)

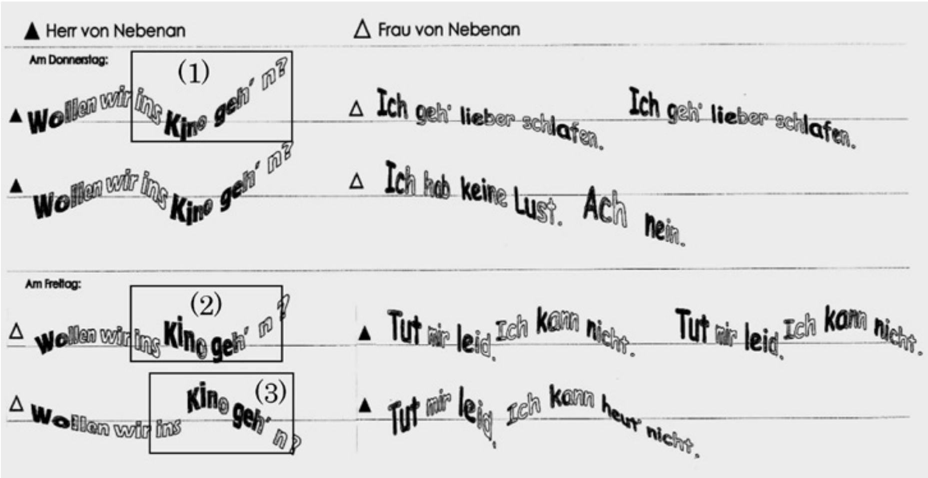

Figure 3.1: Dialog with three different nuclear contours

"Shall we go to the movie theater?": Students in the experimental groups A and C received the dialog in this form of PW, the control group B received text in standard writing. The three different nuclear contours ("Kino geh'n") were better reproduced in the experimental groups A (PW and audio) and C (only PW), though the control group B had access to audio material, whereas group C had only PW but no audio. (Here: Loudness is shown by gray level, black represents maximum loudness or stress, white represents minimum loudness or unstressed syllables; the black & white pattern therefore represents rhythm. Designed by: 鶴田愛 Megumi Tsuruta)

In general, the sub-group that received both, PW and audio, performed best, followed by the sub-group with PW only, and last was the control group receiving just normal text and audio.

A statistical analysis was not performed since the number of students was not sufficient (smallest group: only 8 students). However, the result indicates that supplying only audio materials to students is not enough for the acquisition of prosody. Some students might need an additional form of teaching, and visuals could be an appropriate choice. (Rude 2011)

Another experiment showed also the significant impact of PW on duration of interjections (e.g. in: "Oh, Ihr Deutsch ist wirklich gut!", Engl.: 'Oh, your German is really good!', significant impact for total duration of interjection 'Oh' and the following pause).

A third experiment showed the impact of PW on the correct production of the nucleus, while a control group often stressed the wrong syllable, probably due to interference from English (e.g. in the German word “zentral”, English: ‘central’) (Rude 2012).

Further experiments have to be performed to confirm these findings.

3.2 PW used in classes at Nagoya University

When PW is used systematically in foreign language education, a fixed set of prosodic features to be visualized and their meaning need to be defined. This section will introduce such a canon of ten features to be taught through PW.

In the academic year 2013, prosody was introduced to 1st year students at Nagoya University in German language beginner classes in such a sytematic way. Fig. 3.2 shows an excerpt from a handout given to students during a short lecture on prosody and PW. The example sentence was taken from a sample of PW, which had already been used in the very first lesson during an informal introduction. On this occasion, many of the ten features had already been explained briefly (e.g. by using gesture and voice to demonstrate the relationship between letter size and stress).

Three differences to previous applications of PW (from section 2) should be pointed out:

(1) Size/loudness variations are now scaled (3 levels: normal/normal, larger/secondary stress, largest/primary stress), though there is some variation due to handwriting on ruled paper. A similar restriction (small number of discrete values) has been applied to height/pitch variation (3 levels: middle/normal, high/high, low/low).

(2) The style employed is a broad prosodic transcription (phonological) rather than a narrow one (phonetical) because of the restriction of features/values, as well as through the principle to ignore certain prosodic phenomena (non-visualization of most secondary stresses or non-phonological pitch movements).

(3) There is one deviation from the non-symbolic characteristic of PW: In the past, length of vowels had been visualized merely by stretching the graphs in x-direction in a *quantitative* way. Now, for vowels of accented syllables, there are two different fonts, one representing long (e.g. “A” in SAGEN), the other short vowels (e.g. “A” in LANGSAM), introducing a *qualitative* visual difference (the alphabet of PW thus consists of 38 letters, since all 8 vowel letters of the German alphabet (including 3 “Umlaute”) have now 2 versions, long and short).

Since the 38 letters can be mapped unambiguously onto the 30 letters of the German alphabet, orthography is still obeyed, though upper/ und lower case letters are not distinguished.

In the style used in this class, you can see

(Prosodic feature)	(visual feature)	(possible meaning)
① main stress[2]	large letter(s)[3]	primary importance
② secondary stress	medium-size letter(s)	secondary importance
③ high pitch accent[4]	slightly higher letter(s)	new information
④ low pitch accent	slightly lower letter(s)	old/inferable information
⑤ rising intonation[5]	rising letter string	incompleteness
⑥ falling intonation	falling letter string	completeness
⑦ long stress vowel	1 stretched letter / 2 letters	distinctiveness
⑧ short stress vowel	1 standard letter	distinctiveness
⑨ pause	larger horizontal space	syntactic border; breath
⑩ rhythmic stress	bold letter(s)	establishes rhythm

Figure 3.2: Style of PW used in German classes at Nagoya University
List of ten prosodic features (names were simplified for students; short explanations in footnotes, here omitted), their visual correlates and meaning as explained in lessons. Nuclear/ prenuclear stresses (1/2), associated pitch (high/low: 3/4) and duration (long/short vowel: 7/8); sentence final intonation (rising/falling: 5/6) and pausing (9); plus non-phonological rhythmic stress (10). Example: "Ich bin Ausländer [I am a foreigner] und spreche nicht gut Deutsch [and don't speak German well]. Ich versteh' nicht [I don't understand], was Sie sagen [what you say/are saying]! Bitte [please] langsam [slowly]!"

Experiments have been conducted and are now analyzed to check the impact of this style of PW.

4. Introducing the concept of spatial letters for PW

Up to now, 2-dimensionality of PW was achieved by building a string of de facto 2D letters and by positioning them in 3D geometric-prosodic space. Now, the letters by themselves will become three dimensional. The usage of 3D fonts or 3D typography is widespread in advertisement, and computers make it fairly easy to generate such fonts. However, as with PW, the idea conveyed in this paper is to assign meaning to the 3-dimensionality of letters. Why and how this can be done is explained in the following section.

4.1 Motivation: Why should PW have a 3D font?

If geometrical properties can carry linguistic or paralinguistic information, and if such information ought to be extracted and made use of by reader-viewers, it is only straightforward not to limit the scene to a string of 2D letters curving in 3D space, but to form the letters themselves truly three-dimensionally.

Main arguments for 3D fonts

Human cognition can easily recognize a solid object's position and orientation in 3D space: visual perception has been optimized to perceive, recognize, categorize and track 3D objects, from prehistorical tasks of avoiding predators and catching prey up to today's catching a baseball travelling at 100 km/h.

Visual depth – for example – is a universal feature extractable by all human beings unlike any symbolic representation, which is always culturally biased: consider Japanese students who rely on Katakana for some time, before they master Roman characters for efficiently coding sentences in a Western foreign language.

Purpose of 3D fonts

There are many more features inherent in 3D compared to 2D objects that can serve as additional visual correlates for coding additional phonetic or other features.

On the visual side: 3D ligatures, blends of letters from different alphabets like Roman/Katakana, IPA/Katakana, etc. On the phonetic side: syllabic clustering, segmental features, paralinguistic features like breathyness, lexical features like word class, syntactical features like grammatical case, etc. Depth of the letter – for example – yields an extra metric dimension, the fourth geometric dimension, which can be used to encode additional data into the geometry of individual graphemes. As for PW, the elevation of the letter string (the z-value), and the "thickness" of each letter, (Δz-extension) can express values of different features; z could express "loudness", Δz could express the phonetic feature of "tension" (Sections 4.4 and 4.5).

Quick extraction of relevant visual features of 3D bodies is only one side of the coin. A big advantage is also the ease with which our brain suppresses momentarily irrelevant features, e.g. perspective distortion, shadows, reflections or even partial occlusion.

Hopefully, these general considerations will become more clear in the next section through some illustrations.

4.2 Using stereograms for illustrating the concept

Stereograms are image pairs that – if viewed in the correct manner or by using some special device – fuse into one image and appear to have depth. The perceptual mechanism in short: Basically, humans cannot really „see" 3D images, they can only perceive two slightly different 2D images. After some preprocessing, this pair of retinal 2D images is processed in the brain, which calculates depth information from small differences between the left and right retinal image (and from various other information like color, reflections, shading, etc., not considered in this paper).

Seeing the world "feels" like looking out from the eyes as through two windows in our skull; however, what humans actually do is to re-construct a 3D representation of the world in the brain and to believe and act as if this visual construct would be the real world per se.

So-called "optical illusions" present obvious contradictions between these constructs and reality out there, and thus prove the illusionary character of the visual construct. As often pointed out, the expression is ill-termed, since we always see an illusion. A more appropriate term would therefore be "obviously reality-contradicting optical illusion" opposed to "sufficiently reality-conforming optical illusion".

The small differences between the two retinal images originate from the fact, that our eyes are located horizontally apart from each other and thus perceive slightly different images from objects. The image pair of stereograms contain similar small differences in otherwise almost identical scenes, which are interpreted by our brain as depth phenomena.

Fig. 4.1 shows a simple stereogram that should be viewed in parallel view (e.g. by wearing reading glasses, moving very close, gazing at the black dot and slowly moving away while trying to focus the complete scene).

Be careful and avoid staring at the images for extended periods of time.

Tiredness or even a headache might result.

This simple way of perceiving depth shall just help to understand the following concept of 3D fonts in this paper. In real applications, 3D computer displays or 3D projectors and screens can be applied.

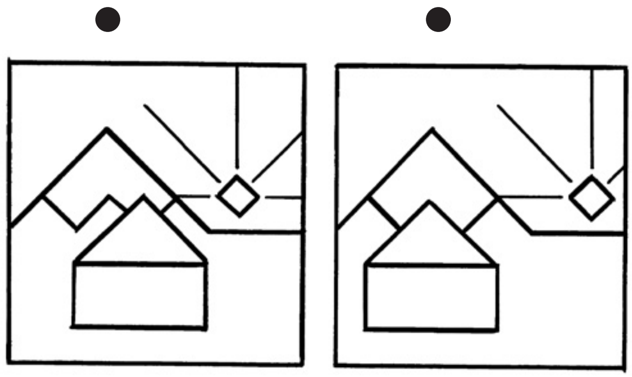

Figure 4.1: A simple stereogram for parallel view.

Stereo pair: Fuji-san at base (stereo distance: 50 mm), house in front, sun behind (3 stacked 2D-images). The parallax (horizontal displacement relative to base stereo point pair's distance) is too big (-5 mm for the house, +5 mm for the sun), therefore, only one of the three objects can be perceived clearly at any given time, the other two appear as double images.

Having done this brief training and being able to see three different layers of "depth" in the above figure (house in front of Fuji san in front of sun), we are ready to move on with the concept.

4.3 Adding perceivable depth to graphs for encoding loudness

In Fig. 4.2, the greeting "Guten Abend, Elke Eber!" is shown in PW in the form of a stereogram. As in Fig. 1.4, the visual axes x, y and z express the prosodic dimensions time, pitch and loudness. Whereas Fig. 1.4 showed only "pseudo-depth" expressed by augmented graphemes (U, A, E), human viewers capable of stereoscopic viewing can actually perceive depth in Fig. 4.2. The four stereograms show – from top to bottom – stronger depth values.

4.4 Adding depth to graphs for encoding tension of speech sounds

In the next figure, all graphemes expressing vowels ("Guten Abend Elke Eber")

Figure 4.2: Using the z-axis (perceivable depth) for loudness.
The same example of PW as in Fig. 1.4 is shown, however, now in the form of a stereogram. Curvature in the vertical (up-down movement on the paper) expresses intonation (pitch movement: "Abend" is high, "Eber" starts low, moves high), curvature in depth (towards the observer) expresses stress, with loudness being the stress correlate (U, A, E(BER) appear closer, sound louder). Top to bottom: Increasing parallax produces increasing depth impressions.

are shown in a simple 3D font, by connected replications of the simple shape, differing just in the z-coordinate. Again, the depth-values increase from top to bottom.

Japanese learners of German might favor this "tension code": They might require a clear discriminatory visualization of tension of E-sounds (tense as in *Eber*, lax as in *Elke*, or reduced to a Schwa as in *Elke*, *guten* and *Abend*, or omitted as possible for the latter two). This could easily be coded by choosing three different depth values for the E-graph in its spatial font in PW (and possibly a "zero"-depth version for a Schwa that can be omitted, as in "guten" and "Abend"). Figure 4.3 shows four such different "allographs" for E (variations of graphs representing the same grapheme E). The transition to the single-valued German grapheme "e" is smooth, given the similarity of shapes differing only in depth. On the symbolic side, learners would have to learn the set of quite different IPA characters [e], [ɛ], and [ə]). In Table 2, the visual correlates of prosodic features as used in Fig. 4.3 are summarized.

Instead of tension, any other meaningful scalable property could be encoded as letter depth.

Figure 4.3: Adding the Δz-dimension (depth of letter) for showing tension
Letter depth encodes the phonetic feature of "tension": Long vowels are tense ("U", "A" and "E" in "Eber") and receive the highest degree of depth, short vowels are lax and have less tension ("E" in "Elke"), reduced vowels (Schwa, like "e" in English "the") have the lowest tension ("e" in "Abend", etc.). Top to bottom: Increasing overall depth values.

Table 2: Summary of prosodic dimensions (left); and corresponding visual correlates or dimensions (right) according to Fig. 4.3.

time (t)	x, horizontal
pitch (P)	y, vertical
loudness (L)	z, depth of letter string
tension (T)	Δz, letter depth
Summary: prosodic feature t P L T	Dimension: 1 2 3 4 visual feature x y z Δz

4.5 Alternative: using depth of graphs for removing vertical curvature

Another option is to alter the prosodic-visual correspondences as in Fig. 4.4, top two, and Table 3: Here, tension is not represented, but time, pitch and loudness are assigned to x, z and Δz, respectively. Henceforth, the up-and-down curvature of letters in the vertical would disappear and be replaced by a back-and-forth curvature in depth.

Note in Fig. 4.3 and 4.4 the typical combination of utterances in German language of high pitch/high loudness followed by low pitch/high loudness in stresses, as in "Abend" and "Eber". Showing pitch/loudness values as two separate visual phenomena that can clearly be discriminated might help Japanese learners to acquire low pitch accents, which do not exist in Japanese.

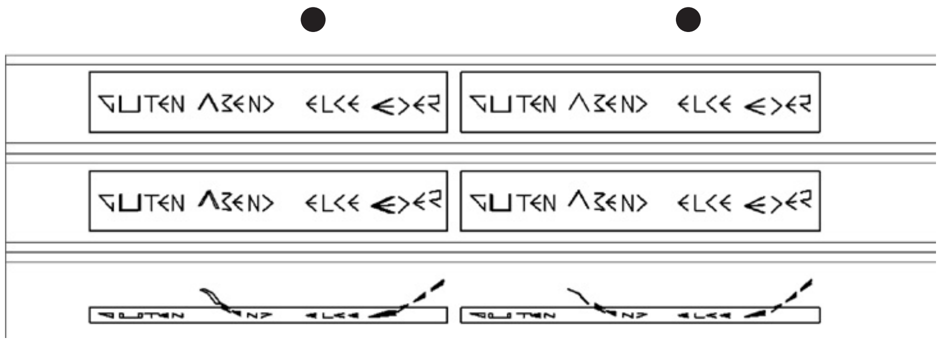

Figure 4.4: Expressing t, P and L by x, z and Δz, respectively; in top view (upper two) and in bird's-eye-view (lower stereogram). The y-axis (vertical on paper) is unused; instead, the z-axis (perpendicular to paper) expresses pitch.

Table 3: Assignment of prosodic and visual dimensions in Fig. 4.4.

time (t)	x, horizontal
×	y, vertical (unused)
pitch (P)	z, depth, height of string
loudness (L)	Δz, character depth
Summary: prosodic feature T × P L	Dimension: 1 × 2 3 x × z Δz

As for Fig. 4.4, text book or e-book designers might favor such an option, since avoiding curvature in the y-dimension (up-and-down curvature of the letter string) saves printing space or space on the display; such a form is also most similar to standard texts and might therefore be preferable for educators. In fact, looking at just one image is almost like normal text. The prosodic features of pitch movement and loudness only pop up in stereographic view.

However, preferences of designers or educators should not be decisive: rather the needs of the learners should be guiding the development cycles of PW and its fonts. The learners' needs are very individual and could require the visualization of some prosodic features in an exaggerated way, of others only very subtly or not at all.

5. Discussion

The starting point of this paper was the research question for an optimum script for language learners. A writing system, prosodic writing (PW), was introduced that encodes also pronunciation features in a given text while preserving orthography (section 1).

The basic form of PW includes a code of the three prosodic dimensions time, pitch and loudness. These dimensions constitute most prosodic features – except articulatory tension. The concept has been described and illustrated in section 2 by using simple shapes for characters.

Section 3: Results achieved in recent years – impact on intonation and durational phenomena – have also been briefly reported, but these results need to be confirmed in further experiments. The system now being applied in classes at Nagoya University encodes ten prosodic features. Notably, its character is now more phonological (broad prosodic transcription) than previous samples (relatively narrow prosodic transcription, e.g. Fig. 3.1), as it deviates from the horizontal central axis only from the nucleus until the end of the intonational phrase (high or low); for prenuclear tones it deviates only slightly to indicate learners high or low pitch accent (see Fig. 3.2).

Recently, the theoretical concept of PW has undergone refinement and distinguishes between flat 2D and spatial 3D fonts. Spatial fonts allow additional degrees of freedom, e.g. the additional dimension of depth of the characters themselves (Δz-values) in addition to the altitude of the letter string above the virtual background (z-values). This fourth dimension can be used to code additional features, e.g. the segmental feature of articulatory tension (see Fig. 4.3), or to release some previously used visual dimensions, for example the vertical curvature in order to allow denser line spacing (Fig. 4.4).

This paper used intentionally simple forms for characters. The question is not one of beauty but of learning efficiency and readability; form should follow functional requirements. Readability and legibility of PW including 3D fonts are essential; however, they have to be redefined in terms of needs and judgments of language learners, but neither by native speakers nor by traditional typographical rules alone.

To return to the research question above: What is the optimum visual code of (spoken) language for language learners? We are still far from an answer: This optimum will look very different for learner groups and even for individual learners; Japanese learners of German or English, for example, might require a clear indication of tension or laxness of E-sounds at the beginning of their language studies (section 4.4).

3D fonts mean much more than just adding depth to each character: It means to look at spatial bodies instead of surfaces with all their potential differentiations. 3D fonts thus allow blends of characters, which appear from one side as one character, from another side as another character. Roman characters (e.g. "E") could be fused with IPA-characters (e.g. "e, ɛ, ə), to show the three different sound-values in side view, while showing the orthographic value in front view. Viewed from in between, a 3D object would be visible that could be interpreted or read as both.

Perspective requires the concepts of allographs and ligatures to be redefined for fonts that are truly 3D.

While all these ideas might seem to be far ahead to some readers, 3D screens are already on the market (Dell, Toshiba and others), some big companies sell more E-Books than paper books, and it is only a matter of time until virtually every computer will have a 3D display. As 3D fonts are already on the market and used in advertisements, pedagogy just has to apply these two 3D techniques for its purposes that can go far beyond our imagination today.

As for PW, teaching materials will – for the time being – make use of 2D representations (see Fig. 5.1). Still, many experiments are to be done using conventional PW (2D characters floating in conceptual 3D space, projected back to 2D). But basic research on PW with 3D fonts should go ahead and can already now help to produce better 2D teaching materials, since if they are proper projections of 3D shapes, our mind is able to reconstruct them in 3D.

Figure 5.1: PW of the phrase above in the style being applied to Nagoya University classes for the time being.

6. Outline

Teaching language by using 3D fonts and writing systems with text floating in 3D is only the tip of an iceberg. "The phenomenal world precedes cognition which precedes language which in turn precedes literacy" (Coulmas 2002, p. 3). However, writing is more: As conventional reading and writing changed in a kind of washback effect the way we think, also reading a 3D text can have tremendous consequences, since it stimulates different parts of the brain simultaneously. We might often prefer to read conventional text, in particular when we read a novel: we like to enjoy the images that evolve from our imaginary power stimulated by a highly abstract text interacting with memory traces of our own experiences. However, there will be plenty of applications for 3D fonts, and language learning is only one of them.

7. Literature

Altmann, Hans and Ute Ziegenhein, “Prüfungswissen Phonetik, Phonologie und Graphemik”, Vandenhoeck & Ruprecht UTB, 2010.

Coulmas, Florian, “Writing Systems. An Introduction to their Linguistic Analysis” Cambridge University Press, 2002.

Gilbert, J. B., “Clear speech from the start: basic pronunciation and listening comprehension in North American English”, Cambridge University Press, 2001/2009.

Haralambous, Yannis. “Fonts & Encodings”, O’Reilly Media, 2007.

Rude, Markus, “Versuch einer Bestandsaufnahme – die nukleare Kontur im Deutschen und ihre Vermittlung an japanische Deutschlernende”. In: Mapping zwischen Syntax, Prosodie und Informationsstruktur. München 2011 (IUDICIUM), p. 55-69. (in German).

Rude, Markus, “Native-like Duration Ratio of Stressed vs. Unstressed Syllables through Visualizing Prosody”. In: Proceedings of Speech Prosody 2012 6th International Conference (Shanghai).