Phonetics in Functional Discourse Grammar[1]

Klaas Seinhorst

Amsterdam Center for Language and Communication, University of Amsterdam[2]

Abstract: The theory of Functional Discourse Grammar assumes that the structure of linguistic items is to a large extent determined by their use, but this idea has not yet been fleshed out in the realm of phonology and phonetics. The present paper aims to address this void, and argues that FDG should be extended with a fifth level of representation, namely a Phonetic Level. With this addition, the model can capture the interaction of two competing functional pressures: the maximization of perceptual distinctiveness, and the minimization of articulatory effort.

1 Introduction

One of the basic assumptions of the theoretical framework of Functional Discourse Grammar (hereafter FDG; Hengeveld & Mackenzie 2008) is the idea that the structure of linguistic utterances is shaped by the demands and constraints that language users pose on them. In the field of phonology and phonetics, this idea has received ample attention since the late 19th century (Passy 1890; De Groot 1931; Trubetzkoy 1939; Jakobson 1941; Martinet 1955, 1961; Stampe 1973; Lindblom 1990; Archangeli & Pulleyblank 1994; Boersma 1998; Kirchner 1998; Boersma & Hamann 2008, and many others); however, functionalist approaches to phonology and phonetics have apparently not received any attention in FDG, in spite of their shared basic assumptions. In this paper, I explore a coupling of the two, and I will contend that it is desirable to extend FDG with a fifth level of representation, namely a Phonetic Level.

This paper is structured as follows: §2 sets out the foundations of functionalism in phonology and phonetics, and evaluates the predictions it makes about speech production, phonological typology and language change; §3 discusses the current role of phonology and phonetics in FDG, and presents the proposed addition of a Phonetic Level; the conclusion remains for §4.

---

[2] Correspondence: Spuistraat 210, 1012 VT Amsterdam, the Netherlands; seinhorst@uva.nl.

2 Functionalism in phonology and phonetics

Phonetics is the branch of linguistics that studies the physical properties of the human speech signal (acoustic phonetics), and the way this signal is perceived and produced by language users (auditory and articulatory phonetics, respectively). It is distinct from – but closely tied to – phonology, the linguistic subdiscipline that studies sounds as discrete, distinctive elements, as well as their internal structure, the way in which they can be combined into larger prosodic units (such as syllables, phonological words and intonational phrases), and their organization into a sound system.

Diverging views exist on the relation between phonology and phonetics. The distinction between the fields was pursued by adherents of the Prague School (e.g. De Groot 1931; Trubetzkoy 1939; Jakobson 1941), who were interested in such notions as the phonological feature, segments and distinctivity. In the generative literature, the distinction was obscured again: for instance, Chomsky & Halle (1968: 5) suppose that a phonetic representation is a "sequence of discrete segments", and nowadays many researchers employ representations that contain phonological as well as phonetic content (e.g. Kirchner 1998; Flemming 1995, 2011).

2.1 THEORETICAL FOUNDATIONS

In functionalist theories of phonology and phonetics it is assumed that the implementation of an utterance is subject to two competing forces: the maximization of perceptual clarity, and the minimization of articulatory effort.

The maximization of perceptual clarity involves the selection of the least ambiguous auditory information pertaining to a phonological structure, i.e. the cues that are least likely to be confused with any cues that signal another meaning. This lack of ambiguity is obviously beneficial in establishing successful communication. Such unambiguous cues tend to occupy peripheral locations in the auditory and articulatory space: for instance, in order to produce an unmistakable [a], signalled by a very high first formant (a resonance in the vocal tract), a speaker needs to lower his mandible considerably. At the same time, however, this speaker wants to minimize articulatory effort, because he strives to execute a minimal number of muscle movements and invest as little effort as possible in them (Zipf 1935, 1949). A reduction of articulatory effort can result, for instance, in the shortening or deletion of phonetic and phonological

content, in a decrease of pitch and peak amplitudes, in the centralization of vowel formant frequencies (resulting in a more schwa-like sound), and/or in tonal sandhi.

Functionalist phonology and phonetics thus predicts that speakers invest only as much articulatory effort as needed to convey their intentions to their interlocutors. This principle is sometimes taken to be listener-oriented, that is, teleological: in this view, the speaker produces an utterance that he judges most likely to be correctly reconstrued by the listener, so he needs to compute a listener-optimal form. A more economical approach postulates that the speaker does not compute such a form, but instead prefers auditory information in production of which he has learnt in perception that it is the least confusable (cf. Boersma & Hamann 2008). Such an approach assumes bidirectional use of linguistic knowledge in the speaker-listener.

2.2 EMPIRICAL EVIDENCE

This section explores some sources of evidence for the predictions made by functionalist phonology and phonetics.

2.2a Reduced speech

Speakers make abundant use of reduced pronunciation variants, such as [hlɛɾɛs] for *hilarious* or [ɔnɹi] for *ordinary*, especially in spontaneous speech. Several experiments have investigated the acoustic properties of reduced speech, and its production and comprehension in the individual. Speakers reduce words and constructions that are highly frequent (cf. the Probabilistic Reduction Hypothesis by Jurafsky et al. 2001); for instance, English *don't* is reduced most substantially – to mere schwa – in the construction in which it occurs most often, namely *I don't know* (Scheibman 2000).

However, frequency of occurrence is not the only relevant factor, as within a stretch of discourse, second instances of low-frequency items are reduced as well (Bolinger 1963, 1981; Fowler & Housum 1987). In the comprehension of reduced speech, context plays a crucial role: reduced forms presented in isolation are harder to recognize than reduced forms presented in context (e.g. Ernestus, Baayen & Schreuder 2002; Zimmerer 2009; Janse & Ernestus 2011). The degree of reduction of an item thus seems to be related to its predictability: more predictable items – either within an entire language or within a stretch of discourse – are reduced more by speakers, and without the help of context listeners experience serious difficulty to understand them.

2.2b Auditory dispersion in phoneme inventories

Functionalist phonology and phonetics also makes typological predictions about the auditory dispersion in segment inventories, i.e. the distribution of the auditory correlates of phonological categories. These inventories are expected to be structured in a way that maximizes between-category contrast while avoiding articulatorily effortful regions. Indeed, this appears to hold true (Flemming 1995; Padgett 2001, 2003; Boersma & Hamann 2008). Along a single auditory continuum, for instance, categories are usually dispersed as shown below. Figure 1 takes as an example the first formant (F_1) in vowels, the auditory correlate of vowel height, displaying common distributions of one through four categories (very few languages have more than four distinctive vowel heights):

ə ──────────────────────►	Dutch, English in some reduced syllables
ɪ ɐ ──────────────────────►	Tagalog
i 'e' a ──────────────────────►	Polish, Spanish
i e ɛ a ──────────────────────►	Dutch, French

F_1
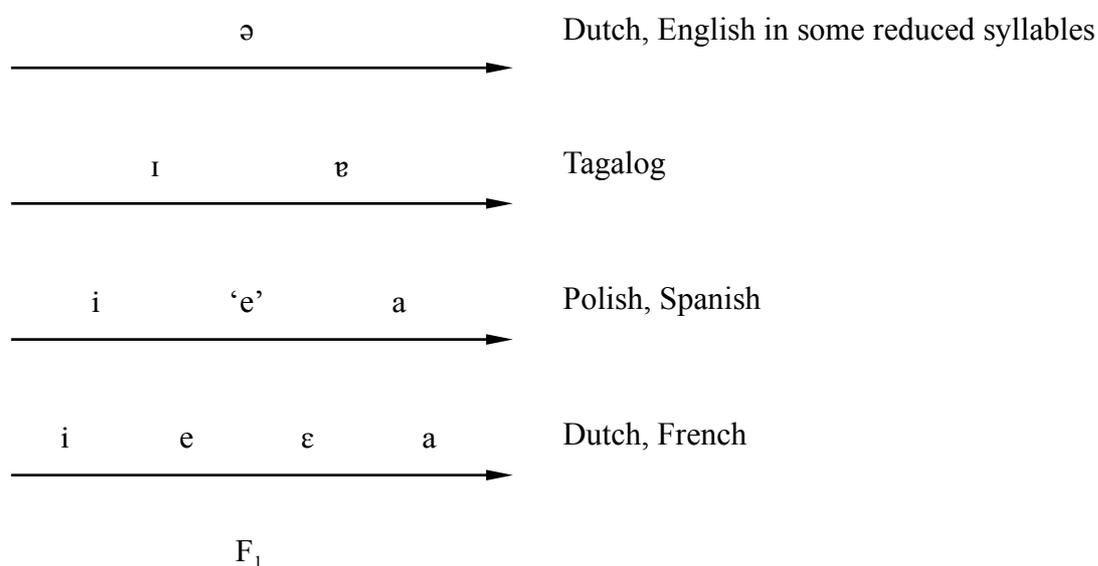
Figure 1. *Distributions of vowel categories along the F_1 continuum.*[3]

If a language has only one category on an auditory continuum, it tends to lie in the center, where articulatory effort is likely minimal; if a language has two categories, they are spaced around the center, ensuring sufficient contrast while still avoiding effortful (i.e. auditorily peripheral) regions; et cetera. Phonological systems thus strike an optimal balance between perceptual clarity and articulatory ease.[4]

---

[3] Single quotes ' ' indicate that no exact phonetic transcription is available for this sound: for instance, 'e' indicates a vowel whose F_1 is intermediate between [e] and [ɛ].

[4] Inventories that are not optimally dispersed will probably undergo sound change. However, speech sounds can be characterized on several auditory continua, and optimal dispersion on one continuum may lead to suboptimal dispersion on another; one sound change thus feeds the next.

2.2c Language change

From a diachronic perspective, it has been postulated that phonological assimilation is the first non-transparent feature that languages exhibit in creolization (Leufkens 2013), indicating an inclination towards articulatory ease.

A topic well-studied within the framework of FDG is grammaticalization (a.o. Keizer 2007; Olbertz 2007; Grández Ávila 2010), a phenomenon where lexical items come to serve as grammatical items, and grammatical items develop new grammatical uses. It is often marked by phonological and phonetic change: e.g. the number of segments in constructions tends to be reduced, pronunciations become less effortful, tonal sandhi occurs. The umbrella term 'phonetic erosion' is often used for these changes,[5] and erosion is the last step in a chain of four processes (Heine & Narrog 2010), occurring after extension (use in new contexts), semantic bleaching (loss of meaning) and decategorization (loss of morphosyntactic properties). Erosion generally presents itself after the grammaticalizing item rises in frequency (Bybee 2003: 147; Bybee, Perkins & Pagliuca 1994: 8, 19). This can be readily explained in a functionalist framework: a rise in frequency makes an item more predictable, and allows for a more economical phonetic form.

3 Phonology and phonetics in FDG

In FDG, as presented by Hengeveld & Mackenzie (2008), phonology is dedicated a representational level within the Grammatical Component. Phonetics resides in the Output Component, whose task is to convert a phonological representation into a gestural plan (Hengeveld & Mackenzie 2008: 8). The model thus incorporates

---

[5] While Dahl (2004) finds 'erosion' an unsuitable metaphor and prefers the term 'trimming' instead, I object to the adjective 'phonetic', because many authors use it to refer to phonological representations. For instance, Bybee, Perkins & Pagliuca (1994: 107) write that "reduction can be manifested in any of the *segmental or suprasegmental features* of the *phonetic* string" (italics mine) and Heine & Kuteva (2006: 62) speak of "phonetic segments". However, segments and suprasegments, as well as their features, are elements of phonology, not phonetics. More precisely, reduction often affects both the phonetic and the phonological forms of a grammaticalizing item. For instance, in reduced items, the auditory cues may be less peripheral (phonetic), but such items may also have less syllables than their corresponding 'full' forms (phonological), segments may be lacking or substituted (phonological), etc.. The *motivation* for these changes, however, is auditory-articulatory in nature, and thus indeed phonetic.

articulatory phonetics, but no mention is made of gestural economy, and it is unclear where exactly a speaker is supposed to specify the auditory events he wishes to realize.

In order to increase the explanatory power of FDG, I would like to assert that the model should be modified in such a way that it captures the two competing forces in functionalist phonology and phonetics – i.e. the maximization of perceptual clarity and the minimization of articulatory effort – without compromising its basic assumptions.

3.1 EXISTING MODELS

The literature provides us with two explicit phonological models that incorporate articulatory effort: one by Kirchner (1998), and one by Boersma (1998 et seq.).

Kirchner assumes two representations in his model (depicted in Figure 2): an Underlying Form, structured in terms of morphemes, and a Surface Form, structured in terms of segments, syllables, phonological feet et cetera. The latter form contains both phonological and phonetic detail, and is evaluated by articulatory constraints that measure articulatory effort (ART in Fig. 2). The Surface Form should probably also satisfy phonotactic restrictions, enforced by structural constraints (Prince & Smolensky 1993/2004; STRUCT in Fig. 2); the mapping between Underlying Form and Surface Form is governed by faithfulness constraints, acting against the deletion, insertion or substitution of phonological material (McCarthy & Prince 1995; FAITH in Fig. 2).

|Underlying Form|

FAITH

/Surface Form/   STRUCT, ART
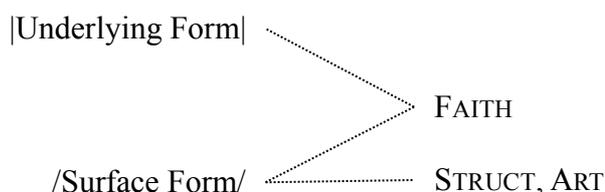
Figure 2. *Kirchner's two-level model.*

Boersma's model of bidirectional phonology and phonetics ('BiPhon'; Figure 3, next page), assumes the same representations as Kirchner's, plus an additional Phonetic Form. This form can be split into an Auditory Form, specifying all auditory events in an utterance, and an Articulatory Form, specifying all muscle activities needed to realize the utterance. Since adult speakers are assumed to have perfect sensorimotor knowledge – i.e. knowledge of the relation between auditory events and articulatory gestures – the Auditory and Articulatory Forms may be conflated onto one Phonetic Form.

In the BiPhon model, phonological and phonetic representations are distinct: the Surface Form contains no phonetic detail. The Surface Form and Phonetic Form are connected to each other by cue constraints (Escudero & Boersma 2004; Boersma 2009; CUE in Fig. 3) linking phonological structure to auditory events: e.g. *[*burst*] /m/ "do not perceive a plosive release burst as /m/" (in perception) or "do not produce an /m/ with a plosive release burst" (in production). The Phonetic Form, or more specifically the Articulatory Form, is evaluated by articulatory constraints: constraints punishing the articulation of more effortful auditory cues are ranked higher than those militating again less effortful auditory cues.

|Underlying Form|

FAITH

/Surface Form/   STRUCT

CUE
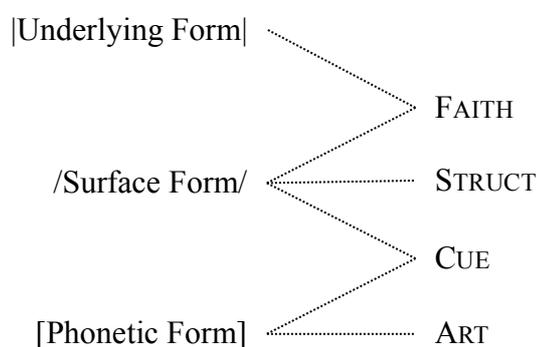
[Phonetic Form]   ART

Figure 3. *Boersma's three-level model.*

The main difference between the models lies in the number of representations: Kirchner assumes two, Boersma assumes three (or actually four, since the Phonetic Form comprises an auditory and an articulatory representation). As a result, different constraint families are involved, and these constraint families evaluate different forms or relations. Because Kirchner's Surface Form contains phonetic detail, it is this representation that must be gesturally economical, and the aim of the faithfulness constraints should probably be twofold: they concern the relation between the morphemic and prosodic content, as well as the relation between morphemic and phonetic material. However, a three-level model appears to comply best with experimental data – for instance, listeners can report their knowledge about the phonological structure of nonsense words, where the lexicon is not involved – and with loanword data, that provide evidence that structural constraints interact with different constraint families in production and perception (Boersma & Hamann 2009).

As far as the embedding in FDG is concerned, Boersma's model seems more suited than Kirchner's, considering the strict separation of the digital and the analogue in FDG (Hengeveld & Mackenzie 2008: 8–9): in the BiPhon model, (digital) phonological forms and (analogue) phonetic forms are not intermixed. I thus propose a fifth level of representation in FDG, namely the Phonetic Level, comparable to Boersma's Phonetic Form. It is to be placed in the Output Component, and can consist of two sublevels: the Auditory-Phonetic Sublevel, specifying all auditory events in the utterance that is to be produced; and the Articulatory-Phonetic Sublevel, containing all muscle commands to instantiate the auditory events from the Auditory-Phonetic Sublevel (cf. Boersma 2011).

3.2 ARCHITECTURE

Figure 4 offers a simple schematization of the proposed revisions of the Phonological Level (O'Neill 2012) and the Output Component (present paper).

Figure 4. *Schematic representation of the proposed revision of the Output Component.*

With O'Neill's (2012) addition to FDG, a distinction between a lexical underlying representation and a prosodic surface representation,[6] the model is able to handle phonological alternations.

As an example of the content of all five representations depicted in Fig. 4, consider the production of the Dutch noun phrase *tien passen* 'ten steps'. This phrase is one morphological phrase, consisting of two morphological words and three morphemes, |tin| 'ten', |pɑs| 'step' and |ən| "PLURAL", so its structure on the Underlying Sublevel could be represented as follows:

(1) $(MP_1$: $(MP_1))$
        $[(MW_1$: $(MW_1))$ $(MW_2$: $(MW_2))]$
            $(M_1$: tin $(M_1))$ $[(M_2$: pɑs $(M_2))$ $(M_3$: ən $(M_3))]$

The abbreviation 'MP' designates a morphological phrase layer, 'MW' a morphological word layer, and 'M' a morpheme layer.[7] Further distinctions within the latter class may have to be made insofar morphosyntactic properties (e.g. root vs. affix) influence phonological properties (e.g. weight, possibility of carrying stress), but I will not make any attempt at that here.

On the Surface Sublevel, no morphophonemic boundaries are present; this representation is prosodically structured, in terms of intonational and phonological phrases, phonological words and feet, syllables and segments. *tien passen* is one phonological phrase, consisting of a single phonological word (assuming that the numeral is unstressed; Booij 1995) and three syllables, so a simplified surface representation could look like (2):

(2) $(PP_1$: $(PP_1))$
        $(PW_1$: $(PW_1))$
            $[(S_1$: tim $(S_1))$ $(S_2$: pɑ $(S_2))$ $(S_3$: sə $(S_3))]$

---

[6] I have adopted a slightly different terminology than O'Neill (2012), who names the representations within the Phonological Level 'Underlying Phonological Level' and 'Surface Phonological Level', respectively. In a similar vein, the proposed representations within the Phonetic Level could also be called 'Auditory-Phonetic Level' and 'Articulatory-Phonetic Level'.

[7] Confusingly, O'Neill (2012) regards the Underlying Phonological Level as a lexical representation (p. 127, 130), but structures it in prosodic terms, identical to the Surface Phonological Level (p. 131). This is not common practice in theoretical phonology, and for O'Neill's analysis it does not seem to be crucial that the Underlying Phonological Level contain prosodic detail.

The three morphemes in (1) and the three syllables in (2) are not isomorphic, due to the syllabification process: if possible, the onset position of a syllable must be filled. Also, the first |n| has been replaced with an /m/ as a result of place assimilation,[8] and the second |n| has been deleted because it is in coda position, following a schwa.

Phonetic forms are by their nature continuous and thus lack internal boundaries. This entails that the notation of the Auditory-Phonetic and Articulatory-Phonetic Sublevels differs fundamentally from that of the levels within the Grammatical Component. There are primitives at the Phonetic Level, namely auditory events and articulatory gestures, but in contrast with the primitives within the Grammatical Component, these have no hierarchical or discrete structure. Representations in terms of these events and gestures are certainly conceivable: they could be multidimensional matrices, reminiscent of Chomsky & Halle's (1968) notations of phonetic forms. However, due to the gradience of the phonetic primitives, their temporal overlap, and the large number of parameters on which they can be described, such representations are quite opaque. I will only provide simple transcriptions in the International Phonetic Alphabet here.

The auditory-phonetic implementation of the phonological surface form (PP_1: (PW_1: [(S_1: tim) (S_2: pɑ) (S_3: sə)])) is [timpɑsə], which would be [[timpɑsə]] at the Articulatory-Phonetic Sublevel.

Given their orthographic similarity, it may seem futile to make a distinction between the two phonetic forms, and perhaps even between both phonetic forms and the phonological surface form, but one and the same alphabet represents fundamentally different entities on the four phonological and phonetic sublevels:

  - in a morphemic representation, a grapheme represents a distinctive unit, i.e. a sound that distinguishes between different meanings – e.g. in Dutch |i| contrasts with |oː|, because the morphemes |tin| *tien* 'ten' and |toːn| *toon* 'tone' refer to different concepts;

---

[8] I assume that |n| is substituted with /m/ on phonetic grounds (the speaker saves himself a tongue tip movement because [m] and [n] are perceptually not very much distinct in this position within the word anyway), but that this substitution has been phonologized, i.e. a phonetically motivated process has become a productive phonological rule. This implies the existence of a Feature layer at the Phonological Surface Sublevel, below the Segment layer. The head of this layer would be an attribute, the operator would be the value of this attribute.

  The alternative would be that the |n| is maintained in the Surface Sublevel. In this case, the /n/ only becomes an [m] in the auditory form, and the phonological surface representation would be (PP_1: (PW_1: [(S_1: tin) (S_2: pɑ) (S_3: sə)])).

- in a prosodic representation, a grapheme represents a bundle of phonological features and feature values – e.g. /i/ is a shortcut notation for a.o. [–consonantal, +open, +tense, +spread, –back];

- in an auditory representation, a grapheme represents a set of auditory events – e.g. [i] is a periodic sound with a fundamental frequency of ca. 180 mel, a first formant of ca. 400 mel and a second formant of ca. 1600 mel (for male speakers; Pols, Tromp & Plomp 1973; Van Son 1993; Wang 2007);[9]

- in an articulatory representation, a grapheme represents a set of articulatory gestures – e.g. [i] indicates contraction of the risoris and buccinator muscles to open and spread the lips, contraction of the masseter and medial pterygoid muscles to raise the mandible, et cetera.

Finally, as the Output I consider the sound waves that result from the implementation of the Articulatory-Phonetic Sublevel. Authors generally use the same notation for the output as they do for the auditory form, but in my view it should be borne in mind that the content of the Output Level is best characterized as purely physical and non-linguistic. What is written as [i] at the Output Level should thus be thought of as the changes in air pressure in Figure 5, showing four periods of this vowel (with amplitude in Pascal on the vertical axis, and time in seconds on the horizontal axis).
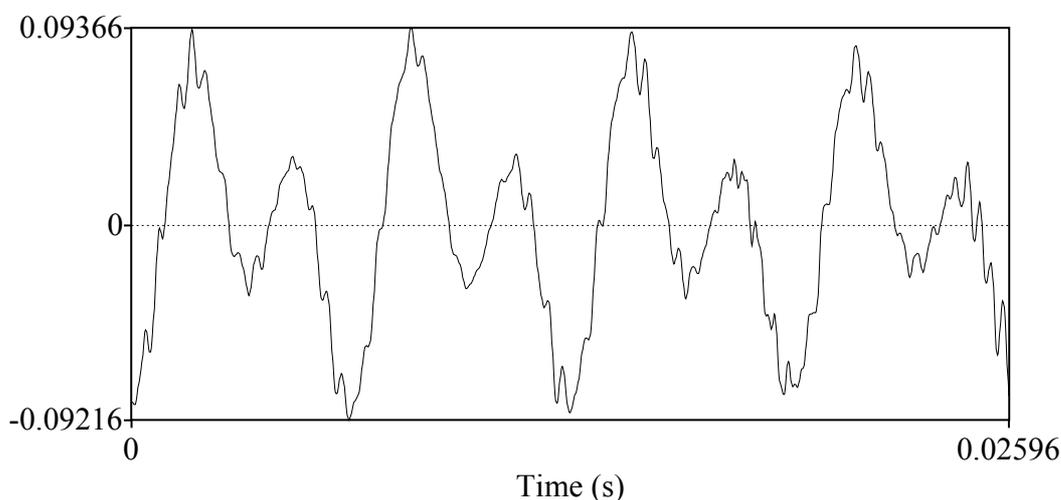
Figure 5. *The content of the Output Level: waveform of* [i].

---

[9] Formant frequencies are often reported in Hertz (Hz), an acoustic scale; the mel scale (Stevens, Volkman & Newman 1937; Stevens & Volkman 1940) is a psychoacoustic scale reflecting the logarithmic nature of auditory perception – on average, humans perceive the pitch interval between tones of 1000 and 2000 Hz as approximately equally large as the interval between 2000 and 3585 Hz (namely 521 mel). Other psychoacoustic measures are the Bark (Zwicker 1961) and ERB (Equivalent Rectangular Bandwidth; Moore & Glasberg 1983) scales.

This waveform is to be described in acoustic rather than auditory terms: within the timeframe of Fig. 5, the sound has a fundamental frequency of 154.1 Hz, a first formant of ca. 300 Hz and a second formant of ca. 2100 Hz. It is the sound waves in the figure that set the listener's tympanic membrane into motion, and that are eventually transformed into a neural excitation pattern. In the listener's auditory cortex, this excitation pattern serves as the input to the language comprehension process.

3.3 Top-down vs. bottom-up processing

In the previous section, I have treated the five levels from Fig. 4 (p. 8) in the order in which they were presented visually in the figure, without having meant to imply that this is the (temporal) order in which speakers compute them – although it is clear that they *realize* the content of the Articulatory-Phonetic Sublevel last, yielding the Output. Note that in Fig. 4, the arrow between the Phonological and Phonetic Levels is double-sided, in contradiction with the basic assumption of FDG that language production proceeds strictly top-down (Hengeveld & Mackenzie 2008: 1–3). According to the principles of functionalist phonology and phonetics, articulatory considerations may pose restrictions on the phonological and auditory forms of an utterance. By means of the Phonetic Level, the Output Component comes to serve as a bottleneck in the speech production process, and exerts a bottom-up influence on it. For instance, it may force the selection of a shorter form that the speaker deems sufficient in the discourse – e.g. *text* instead of *text message*, or *FDG* instead of *Functional Discourse Grammar*; or it may prefer a form with reduced phonetic content and a different phonological structure (one with less syllables, less/substituted segments et cetera). Such alternative or reduced forms indicate that the assumed top-down processing can be countered by pressures of articulatory economy; this is also signalled by the inverse relation between token frequency and token length (a.o. Zipf 1932). Within the model, these pressures go as far as the Phonological Underlying Sublevel; the selection of one of the allomorphs |vjø| vs. |vjɛj| *vieux*/*vieil* 'old-MASC' in French is arguably articulatorily motivated.

This bottom-up influence is also why I have not indicated some sort of "phonetic encoding" process in Fig. 4, which would be more in line with Hengeveld & Mackenzie's (2008: 13) overview of the model; it does not seem to be the case that speakers derive every representation in Fig. 4 from the previous one, but rather that they compute phonetic and phonological representations in parallel (cf. Boersma 2007, 2011;

Boersma & Hamann 2008), thus allowing for a interaction of structural, auditory and articulatory considerations.

This phonological-phonetic implementation of an utterance is clearly influenced by information from the the Contextual Component (although this is not reflected in Fig. 4). This influence is perhaps most obvious in the reduction of speech, where contextual information plays a crucial role both in production and perception; however, speakers also have a large number of articulatory tricks at their disposal to stress their communicative intentions. For instance, they often convey irony – represented at the Interpersonal Level – by lengthening phones, making auditory cues more peripheral than usual (for instance, produce an [s] with a larger amount of hiss (i.e. a higher spectral centre of gravity), or make a velarized l (i.e. [ɫ]) extra 'dark'), producing excessive pitch movements, et cetera. Such exaggerated cues, however, are optional, and the knowledge about their implementation may not be considered to be part of the grammar proper.

4 Conclusion

The present paper intends to bridge the gap between the model of Functional Discourse Grammar and functionalist approaches to phonology and phonetics. The latter share the basic assumption that the phonological and phonetic forms of an utterance are shaped by communicative demands in interaction and physical constraints on the speaker: language users prefer unambiguous auditory cues in order to get their message across clearly (i.e. they strive after perceptual distinctiveness) while investing as little effort as possible in producing them (i.e. they strive after articulatory ease). The consequences of this interaction are visible in several domains, such as reduced speech, the dispersion of categories in segment inventories, and formal change in grammaticalization.

To explain these observations with FDG, the forces of perceptual distinctiveness and articulatory ease should be located somewhere in the model. This can be done with the addition of a fifth level of representation, a Phonetic Level, where the auditory events and articulatory gestures in an utterance are specified (in an Auditory-Phonetic Sublevel and an Articulatory-Phonetic Sublevel respectively; cf. Boersma 2011). Because of the non-discrete nature of its contents, the Phonetic Level is to be placed in

the Output Component. By means of the Phonetic Level, the Output Component functions as a bottleneck in the speech production process, forcing the selection of articulatorily more economical forms that are still perceptually sufficiently salient. This bottom-up influence contrasts with the strict top-down organization of the model as sketched by Hengeveld & Mackenzie (2008).

While FDG does not aim to model language processing in the individual, it is sometimes regarded as a model of the language user (a.o. Boland 2006). The present paper necessarily subscribes to this view, since the driving forces in functionalist phonology and phonetics exist only within the individual. In this perspective, the Phonetic Level is also indispensable to explain language acquisition. Assuming that infants are not innately endowed with language-specific knowledge, they will start building their language system from the two sources of information they have at their disposal: referents in the external world, and speech sounds. Babies perceive the muffled sounds of their future native language(s) already prenatally (mostly pitch contours; DeCasper & Spence 1986; Rosen & Iverson 2007; Mampe et al. 2009), building on their Auditory-Phonetic Sublevel; when they babble, they are exploring the connections between the Auditory- and Articulatory-Phonetic Sublevels.

This paper has only treated spoken language, but its principles apply beyond this modality. For instance, the typology of sign languages is also shaped by the competing forces of distinctivity and economy (a.o. Crasborn 2001; Mathur & Rathmann 2001; Ann 2008; Ormel, Crasborn & Van der Kooij 2013; on grammaticalization: Pfau & Steinbach 2011), and these same principles constrain writing systems as well (Watt 1983; Altmann 2008; Köhler 2008).

References

Altmann, Gabriel (2008). Towards a theory of script. In: Altmann, Gabriel & Fan Fenxiang (eds.), *Analyses of script: properties of characters and writing systems*, 149–164. Berlin/New York: Mouton de Gruyter.

Ann, Jean (2008). *Frequency of occurrence and ease of articulation of sign language handshapes: the Taiwanese example*. Washington, DC: Gallaudet University Press.

Archangeli, Diana & Douglas Pulleyblank (1994). *Grounded Phonology*. Cambridge, MA: MIT Press.

Boersma, Paul (1998). *Functional Phonology: formalizing the interactions between articulatory and perceptual drives*. PhD thesis, University of Amsterdam. The Hague: Holland Academic Graphics.

Boersma, Paul (2007). Some listener-oriented accounts of *h*-aspiré in French. *Lingua* 117: 1989–2054.

Boersma, Paul (2009). Cue constraints and their interactions in phonological perception and production. In: Boersma, Paul & Silke Hamann (eds.), *Phonology in perception*, 55–110. Berlin: Mouton de Gruyter.

Boersma, Paul (2011). A programme for bidirectional phonology and phonetics and their acquisition and evolution. In: Benz, Anton & Jason Mattausch (eds.), *Bidirectional Optimality Theory*, 33–72. Amsterdam: John Benjamins.

Boersma, Paul & Silke Hamann (2008). The evolution of auditory dispersion in bidirectional constraint grammars. *Phonology* 25: 217–270.

Boersma, Paul & Silke Hamann (2009). Loanword adaptation as first-language phonological perception. In: Calabrese, Andrea & Leo Wetzels (eds.), *Loan phonology*, 11–58. Amsterdam: John Benjamins.

Boland, Annerieke (2006). *Aspect, tense and modality: theory, typology, acquisition*. PhD thesis, University of Amsterdam. Utrecht: LOT.

Bolinger, Dwight (1963). Length, vowel, juncture. *Linguistics* 1: 5–29.

Bolinger, Dwight (1981). *Two kinds of vowels, two kinds of rhythm*. Bloomington: Indiana University Linguistics Club.

Booij, Geert (1995). *The phonology of Dutch*. Oxford: Oxford University Press.

Bybee, Joan (2003). *Mechanisms of change in grammaticization: the role of frequency*. In: Janda, Richard & Brian Joseph (eds.), Handbook of historical linguistics, 602–623. Oxford: Blackwell.

Bybee, Joan, Revere Perkins & William Pagliuca (1994). *The evolution of grammar: tense, aspect and modality in the languages of the world*. Chicago: Chicago University Press.

Chomsky, Noam & Morris Halle (1968). *The sound pattern of English*. New York: Harper & Row.

Crasborn, Onno (2001). *Phonetic implementation of phonological categories in Sign Language of the Netherlands*. PhD thesis, Leiden University. Utrecht: LOT.

Dahl, Östen (2004). *The growth and maintenance of linguistic complexity*. Amsterdam/ Philadelphia: John Benjamins.

DeCasper, Anthony & Melanie Spence (1986). Prenatal maternal speech influences newborns' perception of speech sounds. *Infant Behavior and Development* 9: 133–150.

de Groot, Willem (1931). Phonologie und Phonetik als Funktionswissenschaften. *Travaux du Cercle Linguistique de Prague* 4: 116–147.

Ernestus, Mirjam, Harald Baayen & Rob Schreuder (2002). The recognition of reduced word forms. *Brain and Language* 81: 162–173.

Escudero, Paola & Paul Boersma (2004). Bridging the gap between L2 speech perception research and phonological theory. *Studies in Second Language Acquisition* 26: 551–585.

Flemming, Edward (1995). *Auditory representations in phonology*. PhD thesis, UCLA.

Flemming, Edward (2011). La grammaire de la coarticulation. In: Embarki, Mohamed & Christelle Dodane (eds.), *La coarticulation: des indices à la représentation*, 189–212. Paris: Harmattan.

Fowler, Carol & Jonathan Housum (1987). Talkers' signaling of 'new' and 'old' words in speech and listeners' perception and use of the distinction. *Journal of Memory and Language 26*: 489–504.

Grández Ávila, Magaly (2010). *A functional approach to the subjectification of facultative meaning: the case of* capaz *in American Spanish*. MA thesis, rMA Linguistics, University of Amsterdam.

Heine, Bernd & Tania Kuteva (2006). *The changing languages of Europe*. Oxford: Oxford University Press.

Heine, Bernd & Heiko Narrog (2010). Grammaticalization and linguistic analysis. In: Heine, Bernd & Heiko Narrog (eds.), *The Oxford handbook of linguistic analysis*, 401–423. Oxford: Oxford University Press.

Hengeveld, Kees & J. Lachlan Mackenzie (2008). *Functional Discourse Grammar. A typologically-based theory of language structure*. Oxford: Oxford University Press.

Jakobson, Roman (1941). *Kindersprache, Aphasie und allgemeine Lautgesetze*. Uppsala: Lundquist.

Janse, Esther & Mirjam Ernestus (2011). The roles of bottom-up and top-down information in the recognition of reduced speech: evidence from hearers with normal and impaired hearing. *Journal of Phonetics* 39: 330–343.

Jurafsky, Daniel, Alan Bell, Michelle Gregory & William D. Raymond (2001). Probabilistic relations between words: Evidence from reduction in lexical production. In: Bybee, Joan & Paul Hopper (eds.), *Frequency and the emergence of linguistic structure*, 229–254. Amsterdam: John Benjamins.

Keizer, Evelien (2007). The grammatical-lexical distinction in Functional Discourse Grammar. *Alfa – Revista de Lingüística* 51 (2) (special issue: *Advances in Functional Discourse Grammar*): 35–56.

Kirchner, Robert (1998). *An effort-based approach to consonant lenition*. PhD thesis, UCLA.

Köhler, Reinhard (2008). Quantitative analysis of writing systems: an introduction. In: Altmann, Gabriel & Fan Fenxiang (eds.), *Analyses of script: properties of characters and writing systems*, 3–9. Berlin/New York: Mouton de Gruyter.

Leufkens, Sterre (2013). The transparency of creoles. *Journal of Pidgin and Creole Languages* 28 (2): 323–362.

Lindblom, Björn (1990). Explaining variation: a sketch of the H and H theory. In: Hardcastle, William & Alain Marchal (eds.), *Speech production and speech modelling*, 403–439. Dordrecht: Kluwer Academic.

Mampe, Birgit, Angela Friederici, Anne Christophe & Kathleen Wermke (2009). Newborns' cry melody is shaped by their native language. *Current Biology* 19: 1994–1997.

Martinet, André (1955). *Économie des changements phonétiques: traité de phonologie diachronique*. Berne: Francke.

Martinet, André (1961). *Éléments de linguistique générale.* Paris: Armand Colin.

Mathur, Gaurav & Christian Rathmann (2001). Why not GIVE-US: an articulatory constraint in signed languages. In: Dively, Valerie, Melanie Metzger, Sarah Taub & Anne Marie Baer (eds.), *Signed languages. Discoveries from international research*, 1–25. Washington, DC: Gallaudet University Press.

McCarthy, John & Alan Prince (1995). Faithfulness and reduplicative identity. In: Beckman, Jill, Laura Walsh Dickey & Suzanne Urbanczyk (eds.), *Papers in Optimality Theory*, 249–384. University of Massachusetts Occasional Papers 18. Amherst, MA: Graduate Linguistic Student Association.

Moore, Brian & Brian Glasberg (1983). Suggested formulae for calculating auditory-filter bandwidths and excitation patterns. *Journal of the Acoustical Society of America* 74: 750–753.

Olbertz, Hella (2007). 'Dizque' in Mexican Spanish: the subjectification of reportative meaning. *Rivista di Linguistica* 19 (1): 151–172.

O'Neill, Gareth (2012). *Initial consonant mutation in Irish Gaelic: a Functional Discourse Grammar analysis*. MA thesis, MA General Linguistics, University of Amsterdam.

Ormel, Ellen, Onno Crasborn & Els van der Kooij (2013). Coarticulation of hand height in Sign Language of the Netherlands is affected by contact type. *Journal of Phonetics* 41: 156–171.

Padgett, Jaye (2001). Contrast dispersion and Russian palatalization. In: Hume, Elizabeth & Keith Johnson (eds.), *The role of speech perception in phonology*, 187–218. San Diego: Academic Press.

Padgett, Jaye (2003). Contrast and post-velar fronting in Russian. *Natural Language & Linguistic Theory* 21: 39–87.

Passy, Paul (1890). *Étude sur les changements phonétiques et leur caractères généraux*. Paris: Firmin-Didot.

Pfau, Roland & Markus Steinbach (2011). Grammaticalization in sign languages. In: Narrog, Heiko & Bernd Heine (eds.), *The Oxford handbook of grammaticalization*, 683–695. Oxford: Oxford University Press.

Pols, Louis, Herman Tromp & Reinier Plomp (1973). Frequency analysis of Dutch vowels from 50 male speakers. *The Journal of the Acoustical Society of America* 53 (4): 1093–1101.

Prince, Alan & Paul Smolensky (1993/2004). *Optimality Theory: constraint interaction in Generative Grammar*. Malden, MA: Blackwell.

Rosen, Stuart & Paul Iverson (2007). Constructing adequate non-speech analogues: what is special about speech anyway? *Developmental Science* 10: 165–168.

Scheibman, Joanne (2000). *I dunno*: a usage-based account of the phonological reduction of *don't* in American English conversation. *Journal of Pragmatics* 32: 105–124.

Stampe, David (1973). *A dissertation on Natural Phonology*. PhD thesis, University of Colorado.

Stevens, Stanley & John Volkman (1940). The relation of pitch to frequency: a revised scale. *American Journal of Psychology* 53 (3): 329–353.

Stevens, Stanley, John Volkman & Edwin Newman (1937). A scale for the measurement of the psychological magnitude pitch. *Journal of the Acoustical Society of America* 8 (3): 185–190.

Trubetzkoy, Nikolai (1939). *Grundzüge der Phonologie*. Prague: Jednota Československých Matematiků a Fysiků.

van Son, Rob (1993). *Spectro-temporal features of vowel segments*. PhD thesis, University of Amsterdam.

Wang, Hongyan (2007). *English as a lingua franca: mutual intelligibility of Chinese, Dutch and American speakers of English*. PhD thesis, Leiden University. Utrecht: LOT.

Watt, William (1983). Mode, modality and iconic evolution. In: Borbé, Tasso (ed.), *Semiotics unfolding. Proceedings of the Second Congress of the International Association for Semiotic Studies*, 1543–1550. Berlin/New York/Amsterdam: Mouton.

Zimmerer, Frank (2009). *Reduction in natural speech*. PhD thesis, Goethe University, Frankfurt.

Zipf, George (1932). *Selected studies of the principle of relative frequency in language*. Cambridge, MA: Harvard University Press.

Zipf, George (1935). *The psychobiology of language: an introduction to dynamic philology*. New York: Houghton-Mifflin.

Zipf, George (1949). *Human behavior and the principle of least effort: an introduction to human ecology*. Cambridge, MA: Addison-Wesley.

Zwicker, Eberhard (1961). Subdivision of the audible frequency range into critical bands. *Journal of the Acoustical Society of America* 33 (2): 248.