

# Words and Beyond – The Phonetic Channel in Communication

Jürgen Trouvain

Institute of Phonetics, University of the Saarland

## 1. Introduction

The teaching of a language is, to a large part, still based on written language rather than on the spoken language. Thus, very often, questions of *speaking* a language plays the role of "cinderella" in language teaching. However, there is not much disagreement that the most natural way to communicate with one other is through *spoken* communication. In consequence there is a vital need to teach talking and listening in addition to reading and writing (see e.g. Watts, 2000). Though, there is an essential difference between speaking and writing, many people think of sound segments as just spoken forms of letters. So, the first question raised here is: what is the nature of *spoken* words in contrast to *written* words?

A further common assumption is that the verbal content of language represents the core of communication. How communicative information goes far beyond the verbal message in speech is consequently ignored. So, the second question deals with the nature of spoken *words* and what lies beyond the words while speaking.

The second half of the article aims at the implications for language teaching, taking into account the phonetic nature of spoken interaction as the most natural form of communication. Integrating phonetic knowledge and phonetic practice into language teaching covers a wide area. It starts with making learners aware of the importance of phonetic and phonological questions, and ends with the mastering of newly acquired articulation patterns in spontaneous conversation. The conclusion expresses four wishes, addressed to different groups involved in language teaching.

## 2. Basic differences between speech and writing

### 2.1. Manipulating text vs. manipulating speech

In order to show the differences between the basic segments of speech and those of writing the reader is invited to play the following game which s/he may perhaps already know. Let us assume that the basic segments are phonemes in speaking (or a corresponding unit of this size)

and letters and other characters in writing. Do vowels or do consonants contribute more to the understanding of words and sentences?

We manipulate a text – in our case an English sentence - so that we remove all consonant letters, replace them with a space, so that only the vowels are left. The resulting "vowel-only" version of the selected English sentence would look like this:

(1) e ea e oea o o o o : a e ou y i e o i i a e u y e i ea e oo .

Trying to read this will probably fail to get the message.

Now we manipulate the same sentence so that we remove all vowel letters, replace them with a space, so that only the consonants are left. The resulting "consonants-only" version would look like this:

(2) Th w th r f r c st f r t m r r w : r th r cl d n th m r n g w th f w s n n sp lls n th ft r n n .

With some minor difficulties, you will probably cope with reading and understanding it. The complete sentence looks like this:

(3) The weather forecast for tomorrow: rather cloudy in the morning with a few sunny spells in the afternoon.

The summary so far is that in *written* English sentences, vowels bear less information than consonants.

If the assumption is correct that sound segments are roughly spoken letters then we would expect similar results with a speech file manipulation.

Since this article uses the written medium there is no possibility to demonstrate the audio examples which are, of course, at the heart of such an article. It is not very striking to only read the arguments – listening to examples is more convincing. Therefore the reader is referred to the world wide web to get access to audio examples:

<http://www.coli.uni-sb.de/~trouvain/porto.html>

The speech files are as follows. First there is a natural recording of sentence (3) by a female speaker (audio example 1). Second, all portions in this speech file which were identified as representing vowels were replaced by silence using a standard speech editor. The resulting "consonant-only" speech file (audio example 2) consists consequently of silences (the former vowel portions), the consonant portions and the two breath pauses. Third, a similar procedure was applied to get a spoken "vowel-only" version: all portions from the original speech file which were identified as representing consonants were replaced by silence. So, audio example 3 consists of the vowel portions, silences (the former consonant portions), and the two breath pauses. The above mentioned assumption that sounds are spoken letters predicts that the spoken "consonant-only" version is understandable without too many problems, and that the spoken "vowel-only" version is not understandable. However, listening to both audio examples will for most, or even all listeners show the opposite. In the spoken form the "vowel-only" version is understandable with minor difficulties, and the "consonant-only" version is incomprehensible.

How can these results be explained? The first of three phenomena which can explain the just demonstrated effect is *coarticulation*. Letters, characters, but also phonemes are *discrete* units. Sound segments we think to identify in the speech stream are parts of *continuous* articulatory movements. Articulating speech is not articulating sound x, then going back to the resting position, then articulating sound y, going back, and so on. Articulating speech means continuously moving articulating organs. These articulatory gestures overlap in time: before gesture x ends, gesture y has already started. For example before the [m] in "morning" becomes the vowel [ɔ] the velum is still lowered (nasality), the jaw is beginning to lower and move the lips apart (place of articulation) and the tongue and lip shape for the [ɔ] is already there. The time point where in the acoustic waveform the vowel starts is the time point the air is expelled through the mouth. So, consonantal information of the preceding and the following consonants is included in the vowels. For this reason the "vowel-only" version is, in phonetic reality, a "vowel-with-a-bit-of-consonant" version.

The second phenomenon is *timing*. The former consonants which are silences in the spoken "vowel-only" version contain no information about their sound identity anymore. But even if

these silences do not say anything about *which* consonants should fill in these silent gaps they contain information:

- a) the duration of the consonantal stretches encoding the rhythmical structure
- b) the relative total duration of consonants compared to vowels
- c) the temporal location of consonants. This supports the parsing of the syllable structure.

The importance of the timing aspects can be demonstrated with a different manipulation of the original speech file. In audio example 4 the consonant portions are completely deleted, whereas in the "vowel-only" version (audio example 3) the consonant portions are replaced with silences. It can be assumed that the "no-silences" version in audio example 4 is not understandable.

The third and last phenomenon to mention is the *melodic* information. Vowels are usually voiced sounds. That means that the vibration of the vocal folds also bears information on the intonational form. The listener gets information about which words and syllables are marked by a pitch accent and where in the acoustic stream an intonation phrase boundary has been realised. A monotonous realisation of the vowel-only version would not inform us about pitch accents nor about the starts and ends and the types of intonation phrase. The reader/listener can compare audio example 3, the "vowel-only" version, with its monotonous counterpart in audio example 5, in which the fundamental frequency (F0), responsible for the perceived pitch, has been levelled to the average value of this female speaker.

## 2.2. *Connected speech*

To make the differences between speech and writing more distinct some simple statements about writing can help. Writing can be seen as putting one word after another with spaces between them. Some special signs are put at specific between-words spaces according to specific punctuation rules. It is a matter of course that:

- a) a word taken from a sentence context looks the same as written in isolation
- b) this word still looks the same when put at a different location in the sentence
- c) this word looks the same when blanks between the words are deleted

Speech is very different. First, in most cases there are no acoustic "blanks", i.e. silences between words in spoken sentences. Of course there are also pauses in read-aloud speech, but they usually occur at the end of clauses and sentences. If we make a speech file corresponding to the written

sentence, i.e. by inserting acoustic blanks between words would the message still sound acceptable to us?

In audio example 6 you will hear sentence (4) as connected speech recorded with a male speaker.

(4) "The president will be elected for a period of four years."

Two versions with acoustic blanks between words were then created:

i) in audio example 7 the single words were excised from the original connected speech in example 6 and then short pauses were inserted after each word.

ii) all eleven words from sentence (4) were spoken in isolation (audio example 8) and concatenated with longer pauses in one speech file.

Both "blank" versions sound very unnatural and much more attention than normal is required to follow them.

The next manipulation was carried out to show that isolated spoken words are not just words that can be inserted anywhere in a spoken text; they represent whole utterances of their own.

All isolated words were concatenated without any pauses between the words (audio example 9).

The resulting unnaturalness can be partly explained by the fact that the eleven concatenated isolated words are eleven intonation phrases with falling intonation (symbol used here: \) rather than one single intonation phrase with falling intonation:

(5) isolated words: the\ president\ will\ be\ elected\ for\ a\ period\ of\ four\ years\

(6) connected speech: the president will be elected for a period of four years\

Apart from phrasing, the articulation of the *function* words differs a lot between the isolated and the connected form. In audio example 10 all function words or combinations of function words are sampled for direct comparison:

- the special isolated form [ðɪ] for "the" vs. [ðə] in connected speech
- the two words "will be" are [wɪl bi] in isolation but [wəbi] in connection
- the two words "for a" are [fɔ eɪ] in isolation but [fərə] with a linking "r" in context
- and "of" changes from [ɒv] to a reduced [ə] where the [v] has been assimilated to [f] of the following "four"

The audio examples demonstrate that connected speech is very different to the consecutive articulation of words in their lexical form. Sentences as connected speech are very different to sentences in writing.

### 2.3. Dialogues

Communicating with each other normally implies speaking to each other. We take it for granted how we communicate in dialogues (or polylogues). How difficult it really is can be shown by some examples from man-machine interaction.

In dialogues, the interactants must share space and time. Since the invention of the telephone it is possible to communicate by speaking without sharing space. The growing market of mobile telephones shows that there is a great need for such a form of communication. In the last decade or so, it has become popular to communicate by speaking without sharing time and space.

Telephone answering machines got their name from the English expression "to answer a phone" with the meaning of "to lift the telephone receiver". A quick survey in some European languages revealed that the "answer phone" were translated directly into these languages: German "Anrufbeantworter", French "répondeur", Dutch "antwoordapparat", Finnish "puhelinvastaaaja" Swedish "telefonsvarare". The Italian "segretaria telefonica" is even a "telephone secretary". But all these translations bear a gross semantic error: answering machines never *give* an answer - they are just simple voice recorders.

A normal conversation is much more than a pattern of alternating voice messages: message person A – message person B – message person A and so on. Anybody who has tried to have a dialogue by electronic mail knows how long and inflexible, and prone to misunderstandings, this dialogue pattern works. Writing seems per se to be monologic linguistic behaviour, whereas speech is rather dialogically oriented and therefore more flexible in social interaction.

To finish this section with another example from the realm of the man-machine communication, a so called "dialogue system" has been selected. Audio example 11 contains the beginning of a telephone-based dialogue between a caller with a train time table inquiry system. Although this recording is in German which might be an obstacle for the reader/listener, s/he will recognise without understanding the words, how, and perhaps why, the dialogue fails. This example shows the problems of how to model dialogues, and that *words* do not play the most important role.

#### 2.4. *Attitude vs. words*

If we recall the contributions of the caller in the dialogue from the previous audio example (audio example 11) we note that the attitude expressed by his speech changed over time. To demonstrate it with one word, namely the place name "Rosar", he uttered it first in a rather neutral way, the second time with emphasis, and the third time in rage (listen to audio example 12).

This example shows that emotions and attitudes are easily expressed in speech without using explicit words. In most cases even the nuances of the speaker's attitude are decoded by the listener. In "The intonation of American English" (1945: 22) Kenneth Pike brings this everyday experience to the point: „ ... the hearer is frequently more interested in the speaker's attitude than in his words – that is, whether a sentence is ‚spoken with a smile‘ or with a sneer ... “.

Take as another example the following English sentence:

(5) "Could you come to my office?"

Linguistically seen sentence (5) is a question. The prosodic realisation determines whether it is meant as a polite request or a not so polite command (audio example 13). The "tone of voice", i.e. the level and range of pitch, the loudness, the tempo and the voice quality contribute (among other parameters not mentioned here) to the expressiveness of speech, which is missing in writing.

#### 2.5. *Visible speech*

The expression of emotions and attitudes is not restricted to the voice and speech alone. The lips and the jaw, which represent the visible part of the speech articulators, can also be used for facial expression (e.g. smiling, laughing, sadness). Apart from the mouth shape people do other forms of "nonverbal" communication: they move with their heads, they have eye contact with the communication partner, they blink with their eye lids, they frown with their foreheads, they gesture with hands and arms, and they move their whole bodies.

All these visible "articulations" contribute to a conversation as a multi-modal sensation. This can be seen in the co-occurrence of pitch accents with deictic arm and hand gestures, and the temporal correlations between pitch accents and eyebrow movements that can sometimes be observed.

Everybody who has watched a person at a telephone, or people in groups in a restaurant, knows that gesturing is an integral part of communication. Writing, of course, lacks this kind of information.

### *2.6. Most natural form of communication*

To summarise some characteristics of speech which are substantially different to writing: the most natural form of communication makes use of

- a) speech
- b) connected speech,
- c) in dialogues
- d) expressing emotions & attitudes non-verbally
- e) in spontaneous mode
- f) in face-to-face situations.

If the reader/listener agrees with these points then it is only logical to integrate the basics of spoken interaction, namely phonetics and phonology, more thoroughly into language teaching.

## **3. Goals in language teaching**

The proposal followed here how to integrate phonetics & phonology in teaching stems from Barry (1996). He formulated four sub-targets:

- i) accept the need to change pronunciation habits
- ii) understand aspects of pronunciation that need attention
- iii) practice articulatory gestures
- iv) integrate acquired habits into free speech

### *3.1. Accept the need to change pronunciation habits*

Many language learners seem to think that they can master the different pronunciation patterns of the target language *without problems*. They accept that they speak with a "foreign accent" – without considering the possible consequences.

The most serious problem is, of course, that the intelligibility suffers systematically (not to mention the extreme but not so infrequent cases where students are completely unintelligible).

This can be shown with two examples with English as target language: German learners of



English commonly replace the English phoneme /æ/ by /ɛ/; a substitution the students are often aware of. However, they are usually not aware of the final devoicing of plosives and fricatives ("deutsche Auslautverhärtung") in their native language which lead to an ignoring of many phonemic contrasts in the target language. The result is that they may use just one form [bɛt] for four different words: "bet, bed, bat, bad".

Italian learners of English often show substitutions of vowel quantity which causes them to miss some vowel contrasts. As an example: Italian learners realise the English tense and long /i:/ with a short and tense [i] so that "a *sheet* of paper" sounds like "a *shit* of paper" with an Italian accent. Another problem is the realisation of the phonemic contrast of "voicing" in plosives. In English the "unvoiced" plosives /p,t,k/ are aspirated in most contexts in contrast to "voiced" unaspirated plosives /b,d,g/. However, in Italian and other Romance languages /p,t,k/ as well as /b,d,g/ are not aspirated. The sentence "what a shrill *pitch*" would sound like "what a shrill *bitch*".

A "foreign accent" can also have dramatic consequences for social acceptance. A not unpopular British joke aiming at Germans goes like this:

(6) "SS Trooper to his broken watch: 'Ve haf vays off making you tock!'"

To understand the joke one has to know three things: a) clocks in English make "tick-tock", b) British jokes about Germans sometimes refer to the Nazi regime, and in the above situation to a "police" investigation, and c) some characteristics of a German accent: substitution of /w/ by /v/, devoicing of /v/ to /f/ word-finally, and shortening of a long /ɔ:/ to a short [ɔ]. How close at hand those clichés are is shown on the title page of the English mass-tabloid *The Sun* after the resignation of the well-known German federal minister of finance in 1999 (who was disliked by that newspaper). The bold letters covering the whole title page said:

(7) "Ve haf vays off making you QUIT!"

This example demonstrates how quickly and easily national and cultural stereotypes come to the surface, and how a "foreign accent" is an appropriate means for social identification and hence for discrimination.

### 3.2. *Understand aspects of pronunciation that need attention*

In Germany the English language has been the number one foreign language for decades. Despite this undisputed dominance, the pronunciation skills are at a lower level (even at Universities). Pascoe (1996) lists the following five *very common* pronunciation deficits of German pupils of English:

- /æ/ is realised as [ɛ]
- /w/ is realised as [v]
- no distinction between the clear variant and the dark variant of /l/
- missing weak forms: the sentence "Time for you to go." with the function words "for, you, to" in their full lexical form rather than in their reduced Schwa-like form.

In order to avoid pronunciation deficits, it is necessary to *understand* the articulatory gestures of the problematic sounds. This can be done with verbal descriptions of articulatory dimensions. The introduction of phonetic sound symbols is combined with a rather low effort compared to the use for understanding phonetic and phonological processes and elements as well as for using dictionaries for the sake of pronunciation. For reasons of distribution the standard phonetic symbols of the International Phonetic Alphabet (IPA) (see IPA handbook and [www page](#)) should be preferred to other non-standard and idiosyncratic alphabets with low distribution density. Complementary to the use of symbolic information it is important to *show* articulatory gestures. This can be done by looking at visible parts of the speech organs in action: rounding, spreading, protruding of the lips; retracting of the lower lip; degree of opening of the jaw for vowel articulation; place of articulation; dental and alveolar place of articulation of the tip of the tongue. For the less visible, and for the invisible articulations a graphic presentation of the shape of articulations which can be found in most text books on phonetics is necessary. This can help to identify the different manner and places of consonant articulations which is a prerequisite to developing a feeling for the new and the old articulation patterns. These articulatory "gymnastics" on the speech production level need to be accompanied by listening demonstrations and exercises to establish the new patterns on the speech perception level.

In summary, *knowing* about the new contrasts, *feeling* the new articulation and *listening* to the new patterns are the three activities which should go hand in hand to raise the awareness.

### *3.3. Practice articulatory gestures*

The practice of new articulatory gestures should not stop at training the new sounds in isolated words and syllables. Further steps can include nonsense words to focus on the articulatory form rather than the meaning.

The next step should include short phrases of two or three words or morphemes. This size of speech stretch is appropriate to learn more about weak forms. But also rhythmical alternations occur in short prosodic phrases. A very frequently neglected aspect of pronunciation teaching is lexical word stress. Stressing the incorrect syllable in words and short phrases makes the decoding process rather difficult and contributes heavily to the perception of a "foreign accent". The last step should deal with prosodic patterns on the sentence level. It includes the use of *where* and *how* pitch accents are realised, *where* and *how* prosodic phrases are delimited, and on the use of intonational contours.

All these steps have in common that the learners go through the "strange feeling" stage in order to establish new motor and execution patterns on the production side, and new auditory patterns on the perception side. The role of the teacher for delivering training exercises cannot be emphasised enough, to test progress and to give feedback, which is indispensable for the learner.

### *3.4. Integrate acquired habits into free speech*

The last but most difficult sub-target of pronunciation teaching is the integration of the freshly acquired communication habits into spontaneous speech. On the one hand, a speaker's communication goal is to express something meaningful. On the other hand this very speaker has practised the execution of speech in the foreign language by ignoring the meaning. The learner has to monitor two channels: the form of sound function and the form of meaning.

Conversational skills are of course best practised in real situations. If it is the goal of language teaching to master everyday spoken interaction, e.g. to have a political discussion in a pub, then the optimal way for the learner is to practice political discussions in a pub, where the target language is spoken. Going abroad seems inevitable. However, learning in small groups, with peer pressure operating can be one form of preparing for a stay outside the native language community.

#### **4. Concluding wishes**

As a conclusion to the imbalance between the written and the spoken word in language teaching, I present four wishes here which I would have given as answers to a (never raised) question of the fairy in a fairy-tale.

##### *4.1. Pronunciation in dictionaries and teaching material*

The impression of how easily the "foreign accent" is accepted in foreign language learning matches the impression of how half-hearted questions of pronunciation are treated in dictionaries and teaching materials. If pronunciation appears in a dictionary - which is not at all to be taken for granted – then the introductory part of the dictionary, which should be meant as a guide to the use of the phonetic symbols is often inconsistent. The symbols after the lemmata themselves do not always match the proposed guidelines in the introduction. For this reason I direct a wish towards the makers of dictionaries and teaching material that questions of pronunciation should be treated thoroughly, and with the necessary care it requires – or they should give up completely.

##### *4.2. Minimum phonetics and pronunciation lessons in language teaching*

It is remarkable how little is known about pronunciation and phonetics. It takes very little time to present the main deficits and make them known. Spending a few hours to explain some articulatory basics in order to come to the main difficulties of pronunciation seems worth the effort. It makes sense not just to point to differences in the phoneme inventory of both languages, as is usual in contrastive phonology. Explaining phonological phenomena (such as the final devoicing example above), phonetic realisation (such as the aspiration example above), phonemic reductions (such as the weak forms example above), allophonic variation (such as the /l/-variants example above), and especially suprasegmental phenomena such as lexical stress and the location and realisation of pitch accents should be included.

But not only explanation, also correction should be common use in the classroom. Giving learners feedback is a necessary follow-up to making them aware of questions of pronunciation.

##### *4.3. Cross-cultural differences in phonetic communication*

Wish number three is directed to researchers. Doing communication is far more than having a command of linguistic rules. Your linguistic abilities in the foreign languages can be perfect, but

this perfection does not preserve you from a failing to communicate properly in this foreign language. For example, pleasantness in speech is reflected especially by pitch level and pitch range. Reinke (2000: 71) reports a pilot study where Russians and Germans were asked to evaluate audio sequences. Both groups agreed on identifying angry and friendly speech sequences. But they differed regarding the scaling of friendly/joyful sequences: Russians marked a "pleasant" on the pleasant-unpleasant scale whereas for Germans these speech portions sounded "a bit and sometimes very unpleasant". This phenomenon sometimes contribute to cultural stereotypes that Germans are seen as rude and unpolite from a Russian native speaker point of view. Mastering these paralinguistic features relevant for emotions are important for behaving properly in the foreign language community (for further examples see e.g. Reinke, 2000). It is important that in addition to teaching the linguistic competence, the communication competence and the cultural competence is addressed as a goal in language teaching. Although a few studies have started to investigate the field of cross-cultural differences from the phonetic perspective there is enough space and much need to widen this line of research.

#### *4.4. Creating of helpful material and software*

The last wish goes equally to teachers, researchers and software developers. Multimedia applications are *theoretically* ideal for working with authentic text, speech and video material. *Theoretically*, computer-assisted language learning provides an excellent framework for self-learning. However, in practice this means that the offer of computer-assisted pronunciation learning ranges from dull pronunciation drills (but now in digitalised form) to programs which attempt to consider learning relevant factors such as individual difficulties, expected source language deficits, degree of exercise, pronunciation error analysis, differentiated feedback and user-friendly usage (for a review of German pronunciation learning software see Richter, 2000). The technology of automatic speech recognition has not yet reached the level of maturity that garantuees reliable pronunciation feedback. There is some success in automatically rating the fluency of a language learner (e.g. Cucchiarini et al., 2000), but none of the *important* tasks of a teacher can be compensated for with technological aids, namely:

- to provide with his/her own pronunciation as the baseline for the students
- to function as a guide for a correct pronunciation
- to coordinate of the learning route of the students
- to motivate the students

No matter whether with or without technology, the role of the teacher in the acquisition process of new auditory and articulatory patterns is central. A solid understanding of some basic phonetic knowledge, and the most basic is the differences between speech and writing, is an indispensable prerequisite to fulfil this role appropriately.

## **References**

Barry, W.J. (1996), "The relevance of phonetics for pronunciation training", in: *Phonus 2* (Working Papers Phonetics Saarbrücken), pp. 5-19.

Cucchiarini, C., Strik, H. & Boves, L. (2000), "Quantitative assessment of second language learners' fluency by means of automatic speech recognition technology", *Journal of the Acoustical Society of America* Vol. 107 (2), pp. 989-999

*Handbook of the International Phonetic Association. A Guide to the Use of the International Phonetic Alphabet*, (1999), Cambridge University Press.

IPA web page: <http://www2.arts.gla.ac.uk/IPA/ipa.html>

Pascoe, G. (1996), "Pronunciation analysis and teaching", in: *Phonus 2* (Working Papers Phonetics Saarbrücken), pp. 109-122.

Pike, Kenneth (1945), *The intonation of American English*, Ann Arbor.

Reinke, Kerstin (2000), "Ein Babylon der Emotionen? Das Problem der kultur- und sprachenübergreifenden Forschung der phonetischen Emotionssignale", in: *DaF (Deutsch als Fremdsprache)* 37 (2), pp. 62-72.

Richter, Regina (2000), "Selbst gesteuerter Ausspracheerwerb via Multimedia. Lerntheoretische und fertigkeitenbezogene Anforderungen an multimediale Lernumgebungen", in: *DaF (Deutsch als Fremdsprache)* 37 (2), pp. 73-81.

Watts, Sheila (2000), "Teaching talk: should students learn 'real German'?" in: GFL (German as a Foreign Language) 1, pp. 64-82.