Gestalt Psychology Meets Phonetics – An Early Experimental Study of Intrinsic F0 and Intensity

Bernd Möbius

Institute of Natural Language Processing, University of Stuttgart, Germany
bernd.moebius@ims.uni-stuttgart.de

ABSTRACT

Many studies based on instrumental techniques have been carried out since the 1960's to investigate vowel-intrinsic fundamental frequency, intensity and duration. However, the first systematic work on the inherent fundamental frequency and intensity of vowels—in fact, of consonants too—appears to have been performed as a series of perception experiments during the years 1913–1915 by the German experimental psychologist Hans Ruederer; it was published in his doctoral dissertation (Munich, 1916). The results of his perception experiments are compatible with those obtained by instrumental phonetics, and generally far precede them.

1 INTRODUCTION

The study of vowel-intrinsic fundamental frequency (F_0), intensity and duration has received considerable attention in the past decades. Many studies based on instrumental techniques have been carried out since the 1960's. Earlier, the phenomena had been described mostly impressionistically (but cf. [1]).

Vowel-specific F_0, intensity and duration are usually interpreted as microprosodic variations of the speech signal that are due to physiological properties of the human speech production apparatus and to acoustic conditions during speech production. They are therefore often considered to be language-independent and beyond the active control by the speaker. The widely accepted explanation for vowel-intrinsic F_0 (henceforth IF_0), for instance, is a variant of the "tongue pull" hypothesis [2], which assumes a mechanical coupling between the phonatory and articulatory systems. Yet, there is evidence that phonological concerns may override the apparently universal phenomenon, because IF_0 can be constrained in some tone languages [3].

Results obtained from instrumental studies indicate that there is a significant correlation between IF_0 and vowel height or, in acoustic terms, a significant negative correlation between IF_0 and F_1: all else being equal, high vowels have a higher IF_0 than low vowels. There is also a tendency for back vowels to have a higher IF_0 than front vowels. Effects of phonological vowel length (or the tense/lax contrast), syllabic stress, and phrasal position have also been reported. For a detailed review of the pertinent literature the reader is referred elsewhere [4, 5].

As for the intrinsic intensity of vowels, it is generally suggested that, under controlled conditions, open (low) vowels are louder than closed (high) vowels. Vowel-specific intensity is thus correlated with jaw opening and vertical tongue position. One of the suggested explanations for vowel-intrinsic intensity is that the pharyngeal cavity is enlarged during the production of high vowels, causing a stronger dampening effect on the excitation signal for closed vowels than for open vowels [6, 7]. However, the agreement among researchers on the correlation between vowel intensity and degree of openness is not perfect (cf. [8]).

Degree of openness and tongue height are also generally assumed to be the main determinant of the intrinsic duration of vowels. Open vowels tend to have a longer intrinsic duration than closed vowels. One possible explanation is that the jaw must be lowered in a time-consuming gesture to produce an open vowel.

It has been suggested that the relations between vowel height and intrinsic F_0, intensity and duration be considered and investigated as one coherent problem. There is evidence of a complicated interaction of physiological, acoustic and psychoacoustic factors, and the results of this interaction may include compensatory effects in speech production and perception [7, 4].

The first *instrumental* studies of IF_0, and also of coarticulatory or contextually induced F_0 variations, which are often subsumed under the notion of microprosody too, were carried out by means of Meyer's pitch meter and date back to around 1900 (as reported in [1]). The first *perception* study of the inherent F_0 and intensity of vowels (in fact, of consonants too) appears to have been performed during the years 1913–1915 by the German experimental psychologist Hans Ruederer; it was published in his doctoral dissertation [9]. Ruederer also proposed an experimental design for investigating segmental durations but had to accept the fact that these experiments could not be carried out with the technical devices available at his time.

ISBN 1-876346-48-5 © 2003 UAB

2 BIOGRAPHICAL NOTES

Hans Ruederer was born in 1889 in Munich, as the son of a popular local writer, Joseph Ruederer. He studied at the universities of Heidelberg, Berlin, Bonn, and Munich, first attending science classes and later seminars in philosophy, psychology and classical philology. Ruederer was a student of a number of eminent scientists of his time: psychologists (Otto Selz, Oswald Külpe), linguists (Hermann Paul, Karl Bühler), sociologists (Georg Simmel), and physicists (Wilhelm Conrad Röntgen).

The topic of Ruederer's dissertation thesis was suggested and mentored by Karl Bühler, who was then an assistant professor to Oswald Külpe at the Institute of Psychology at the University of Bonn. Külpe, a student of Wilhelm Wundt's in Leipzig, was one of the most influential psychologists of the early 20th century and a co-founder of the Würzburg school of Gestalt Psychology. When Külpe accepted a position at the Institute of Psychology at the University of Munich in 1913, Bühler and Ruederer followed him there.

Ruederer continued his experiments in Munich, but with the start of World War I he joined the voluntary nursing service, which must have left him little time for pursuing his dissertation (as he reports in the preface of his thesis). Evidently, his thesis had originally been designed to consist of three parts, covering the perception of speech sounds, words, and sentences, respectively. Only the first part, however, was completed and eventually submitted, and accepted, as a dissertation thesis (December 1915); it was published in 1916 with the title "On the perception of the spoken word—An experimental-psychological study" (my translation from German) [9]. Ruederer's thesis advisors were Oswald Külpe and Clemens Baeumker.

I have been unable to ascertain what became of Hans Ruederer after 1915, but the fact that the results of his further experiments were never published and the lack of any subsequent publications lead me to presume that he decided not to pursue a scientific career or that the circumstances of the ongoing WWI took the decision away from him.

3 GENERAL METHODOLOGY

Ruederer realized that for the perceptual study of detailed acoustic gestalt qualities, such as speech sound inherent pitch and loudness, utterances produced in real-life communication situations were inappropriate. He concluded that a careful construction of stimulus materials was required. The structure of the stimuli had to exclude the possibility that percepts other than those under investigation were evoked in the listener. Ruederer decided to use isolated sounds as well as nonsense syllables and words as carriers in which the particular speech sounds under investigation were embedded. By using carrier frames he achieved an optimal control of segmental and prosodic contexts: a method that was later employed by many other researchers.

Ruederer had to rely on the human voice to produce the speech stimuli, which called the possibility of identical repetitions of stimuli into question. The only technical devices available at his time for recording and reproducing speech were the phonograph and the grammophone. Both devices were inadequate for Ruederer's purposes, because both modify the original speech signal during recording and while playing them back. During the recordings, the phonograph reduces the intensity of the input signal, albeit not proportionally according to the original spectral energy distributions of different classes of speech sounds, but selectively, such that for instance fricatives are affected much more strongly than sonorant speech sounds. The phonograph has a low-pass filtering effect, and it modifies the spectral properties of the stimuli.

Ruederer acknowledged the fact that, in principle, every technical device for recording and playing back speech distorts the original signal, and concluded: "This is very important and cannot be paid enough attention to, or else one might be of the opinion at the end of an investigation that, based on certain results, one has written a psychology of man, whereas in reality one has established a psychology of the phonograph." ([9], p. 7; my translation).

Ruederer ended up combining both stimulus types, the human voice and phonographic recordings, in his experiments, thereby exploiting each strategy's strengths while avoiding its respective shortcomings. The experiments were aimed at five perceptual correlates of the acoustics of speech sounds: intensity, timbre, pitch, duration, and place/manner of articulation (consonants only). The experiments related to speech sound specific intensity were very elaborate and yielded many detailed results; in this paper I will therefore concentrate on the presentation and discussion of this aspect of Ruederer's work. However, some interesting details of his additional experiments will be addressed too.

4 INTRINSIC INTENSITY

Ruederer used two points of departure for his experiments concerning the perceived intensity of speech sounds. First, he drew a methodological distinction between the acoustically definable intensity of a given speech sound, i.e. intensity as a physical property of speech, and the perceived loudness of the same speech sound. As we know, Ruederer's main interest was in speech perception. Second, he realized that speech sounds differed in terms of their inherent loudness.

C	Rel.Int.	C	Rel.Int.	V	Rel.Int.
h	1.0	ks	4.0	uː	9.1
d	1.8	R	4.6	iː	9.9
b	1.9	w	4.7	yː	10.0
g	2.1	v	4.8	aʊ	10.3
f	2.5	r	4.8	aɪ	10.8
ç	2.7	ŋ	5.0	ɔʏ	11.2
x	2.8	m	5.1	ɛː	11.8
t	3.5	n	5.1	eː	11.9
pf	3.5	l	5.1	øː	11.9
p	3.6	s	5.5	oː	12.1
kv	3.7	ts	5.7	aː	12.5
k^u	3.8	j	6.1		
k^i	3.9	ʃ	6.2		

Table 1: Perceived relative intrinsic intensity of German speech sounds, normalized with reference to /h/. Adapted from Table II in [9], using standard IPA symbols instead of Ruederer's orthographically oriented notation.

Ruederer		Lehiste & Peterson		
V	Rel.Int.	V	Rel.Int.	
uː	1.00	u	(1.00)	80.2
iː	1.09	i	(1.00)	80.2
		ɪ	(1.09)	81.4
		ɚ	(1.09)	81.4
yː	1.10			
aʊ	1.13	aʊ	(1.38)	84.8
aɪ	1.19	aɪ	(1.19)	82.7
		ɛ	(1.21)	82.9
		ʊ	(1.25)	83.4
ɔʏ	1.23	ɔɪ	(1.27)	83.7
ɛː	1.30	æ	(1.22)	83.1
eː	1.31	eɪ	(1.06)	81.1
øː	1.31			
		ɔ	(1.38)	84.8
oː	1.33	oʊ	(1.25)	83.4
		ə	(1.39)	85.0
aː	1.37	aː	(1.44)	85.5

Table 2: Relative intrinsic intensities of vowels. Ruederer data (German) are taken from Table I and normalized with respect to /uː/. Lehiste & Peterson data (American English) are VU meter readings (in dB) for sustained vowels, taken from Table I-A in [6]; corresponding values in parentheses after conversion to the sone scale and normalization with respect to /u/.

The experimental design was as follows. The experiments were carried out in open places, at times when there were no other sources of noise and there was no noticeable wind. The stimuli were produced by a speaker (most of the time Ruederer himself) who had been trained to maintain a constant overall pitch (192 Hz), duration (0.3 s) and volume. Vocalic sounds and continuant consonants were presented in isolation; the text is somewhat ambiguous as to whether stop consonants were presented in isolation too or whether they were embedded in fixed vocalic frames.

The initial distance between speaker and listener was 30 m. After the first presentation of the stimulus set, the listener increased the distance successively in steps of 10 m, along a marked track. At each distance the listener had to identify each of the presented speech sounds. After establishing for each speech sound the distance at which it was no more reliably identifiable, this distance was then reduced in small steps (0.5 m) until the zone was reached where the listener identified the speech sound reliably and correctly. This distance was smallest for /h/, and /h/ subsequently served as a baseline reference.

Table 1 displays the perceived intrinsic intensities of German speech sounds, averaged over all stimulus presentations, and normalized with reference to /h/, as obtained by the method described above. It is striking how neatly the natural classes of speech sounds are grouped together: there is a gap in the otherwise continuous series of relative intensity values that separates the vowels from the consonants. Within the consonants, going from softest to loudest, the voiced stops form a coherent group, followed by non-strident voiceless fricatives, voiceless stops and affricates, sonorants and approximants, and strident fricatives. Within these natural classes, speech sounds are almost "equi-intensive". The vowels are ranked almost perfectly along the openness dimension, with the diphthongs, which have trajectories from open to close, appearing in a somewhat ambiguous position.

This ranking of speech sounds by their perceived relative loudness corresponds well to a ranking by acoustic intensities as obtained by instrumental methods. Table 2 displays Ruederer's results for the vowels along with the well-known Lehiste & Peterson data for American English [6]. In this Table, the Ruederer vowel data are taken from Table 1 and normalized with respect to /uː/. The Lehiste & Peterson data represent VU meter readings, expressed in decibels (dB), for sustained vowels, taken from Table I-A in [6]; the values in parentheses correspond to the original dB values after conversion to the sone scale of relative loudness and a normalization with respect to /u/, for a more convenient comparison. German and American English vowels are presented pairwise wherever feasible.

Even with the qualifications that, first, the small number of vowel pairs makes a numerical correlation analysis not very meaningful and, second, vowel data from two languages may not be straightforwardly comparable anyway, the general compatibility of Ruederer's and Lehiste & Peterson's results is obvious.

ISBN 1-876346-48-5 © 2003 UAB

5 FURTHER EXPERIMENTS

The design of Ruederer's perception experiments on the intrinsic F_0 of vowels (and voiced consonants) is similarly creative as that for intrinsic intensity. For instance, the stimuli were presented by a male speaker who was trained to produce the speech sounds consistently on one of five pitch levels corresponding to the center of the speaker's pitch range (the musical note *g*, or 192 Hz) and the musical notes *c*, *e*, *h*, and *d'*, respectively. Evidently, the intention was to avoid the effects of extraneous factors on F_0, but it may have caused the side effect of constraining F_0 to an undesired extent.

Unfortunately, Ruederer considered the number of data points that he obtained for IF_0 as too small to warrant summarizing them in quantitative terms and presenting them in a table. In his verbal description of the results, an interaction of pitch and spectral energy distribution is suggested. Of particular interest is a passing comment in which Ruederer rejects Helmholtz's conception of the vocal tract as a resonator that selectively enhances the harmonics of the speech signal, and instead advocates an early version of the source-filter model according to which the spectral structure of a vowel is independent of the fundamental frequency of the excitation signal ([9], p. 28).

Constructed nonsense words were used in the experiments on the place and manner of articulation of consonants. The speech sounds under investigation were systematically embedded in the three consonantal positions of CVCVC structures, using several vocalic contexts. The main result of this study is a small set of perceptual cues that subserve a systematic distinction of the consonant classes, viz. sonorants, stops, and fricatives, as well as a finer distinction within these classes according to voicing and place of articulation.

6 CONCLUSION

Lacking the instrumental methods available to researchers half a century later, and indeed aware of the problem that a technical device may distort the original signal, Ruederer put an impressive effort into designing perception experiments that would enable the listener to discern properties of speech sounds as delicate as intrinsic F_0 and intensity.

Ruederer's approach is an excellent example of the meticulous experimental designs developed by the early experimental psychologists. Besides the ingenious spatial arrangement of speaker and listener during the experiments, perhaps the most important design decision was to use carrier syllables and words in which the particular speech sounds under investigation were embedded, thereby achieving carefully controlled segmental and prosodic contexts. Even though Ruederer himself did not apply this method consistently, it was later employed by many other researchers.

The match between Ruederer's results and those obtained by instrumental studies carried out several decades later is remarkable. For instance, Ruederer reports a ranking of speech sounds by their intrinsic intensities that neatly follows the open-close dimension for vowels and the sonority hierarchy for consonants. He also detects systematic differences in vowel-specific F_0, albeit without providing quantitative data.

In summary, the relevance of Ruederer's work to phoneticians is, first, the application of methods developed by experimental psychology to the study of the properties of speech sounds and, second, the fact that the results of his perception studies are compatible with those obtained by instrumental phonetics, and generally far precede them.

ACKNOWLEDGMENTS

I wish to thank Michael Hanke for drawing my attention to Ruederer's dissertation thesis.

REFERENCES

[1] E. A. Meyer, *Die Intonation im Schwedischen I–II*, Stockholm Studies in Scandinavian Philology 10–11. Stockholm University, 1937, 1954.

[2] J. J. Ohala and B. W. Eukel, "Explaining the intrinsic pitch of vowels," in *In Honor of Ilse Lehiste*, R. Channon and L. Shockey, Eds., pp. 207–215. Foris, Dordrecht, 1987.

[3] B. Connell, "Tone languages and the universality of intrinsic $F0$: evidence from Africa," *Journal of Phonetics*, vol. 30, pp. 101–129, 2002.

[4] E. Fischer-Jørgensen, "Intrinsic F_0 in tense and lax vowels with special reference to German," *Phonetica*, vol. 47, pp. 99–140, 1990.

[5] D. H. Whalen and A. G. Levitt, "The universality of intrinsic $F0$ of vowels," *Journal of Phonetics*, vol. 23, pp. 349–366, 1995.

[6] I. Lehiste and G. E. Peterson, "Vowel amplitude and phonemic stress in American English," *Journal of the Acoustical Society of America*, vol. 31, pp. 428–435, 1959.

[7] G. Neweklowsky, "Spezifische Dauer und spezifische Tonhöhe der Vokale," *Phonetica*, vol. 32, pp. 38–60, 1975.

[8] M. Rossi, "L'intensité spécifique des voyelles," in *Proc. 7th Internat. Congr. Phonet. Sci. (Montreal)*, 1971, pp. 574–588.

[9] H. Ruederer, *Über die Wahrnehmung des gesprochenen Wortes. Eine experimentell-psychologische Untersuchung*, Noske, Borna-Leipzig, 1916.