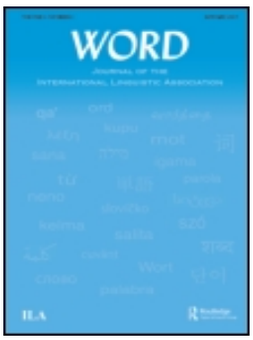

WORD

ISSN: 0043-7956 (Print) 2373-5112 (Online) Journal homepage: https://www.tandfonline.com/loi/rwrd20

Recognition of Word Tones in Whispered Speech

Martin Kloster Jensen

To cite this article: Martin Kloster Jensen (1958) Recognition of Word Tones in Whispered Speech, WORD, 14:2-3, 187-196, DOI: 10.1080/00437956.1958.11659663

To link to this article: https://doi.org/10.1080/00437956.1958.11659663

Published online: 04 Dec 2015.

RECOGNITION OF WORD TONES IN WHISPERED SPEECH

MARTIN KLOSTER JENSEN

The present paper deals with the question of whether, and how, phonemically distinctive tone is rendered in whispered speech. Since it is possible that there may be differences between languages in the manner and effectiveness of whispered tone rendition, the study deals with several "tone languages"—Norwegian, Swedish, Slovenian, and Mandarin—rather than with a single language.

We will first summarize the history of the problem.

The question whether voice pitch is heard in whispered Chinese was brought up by Olof Gjerdman in 1924. Gjerdman had asked a Chinese if two persons could understand each other when whispering in Chinese, and he took the affirmative answer to corroborate his theory that "a whispered language has as many distinct musical accents as the same language when voiced."[1] According to Pike,[2] it is possible that stress replaces pitch in whisper, thus enabling one to distinguish "tones." Pater Giet has maintained that the word tones are both expressed and perceived in whispered Chinese: "Auch beim Flüstern spricht und hört der Chinese die Töne".[3] Eugen Dieth stated briefly that it is possible to indicate intonation in whisper to a certain extent ("einigermassen")[4], and tried to account for this by postulating changes in the air pressure. He tried to invalidate a statement by Gutzmann, that "das Flüstergeräusch ist in seiner Höhe an sich nicht variabel" by referring to the fact that whispered Chinese is fully understandable. This, however, is no proof that pitch exists in whisper: context of course always helps one to understand, and the same may be true of

[1] "Critical Remarks on Intonation Research," *School of Oriental Studies, London Institution*, Vol. III, part III, p. 502.

[2] Kenneth L., Pike, *Tone Languages*, Ann Arbor, 1948, p. 34.

[3] Franz Giet, *Zur Tonität nordchinesischer Mundarten* (Studia Instituti Anthropos), Wien/Freiburg 1950, p. 95.

[4] Eugen Dieth, *Vademekum der Phonetik*, Bern, 1950, p. 96.

stress patterns which, for all we know, may make tone distinctions more or less superfluous.

In 1955, W. Meyer-Eppler made sound spectrograms of a sentence which was first spoken with voice and subsequently whispered.[5] On the basis of these spectrograms he concluded that pitch modulation ("Tonhöhenverlauf") is actually found in whisper, though it is brought about by certain substitutes for phonation. At the same time, G. Panconcelli-Calzia showed that a sentence from a well-known opera text "sung" in whisper to listening subjects was invariably heard as just whispered speech, not "song".[6] Panconcelli-Calzia also had Chinese-speaking subjects listen to whispered "word tones" in Chinese. The way this experiment was conducted was criticized by Pater Giet,[7] who vigorously affirmed his standpoint as expressed in his book.[8]

In 1956, Meyer-Eppler reported on fresh experiments which allegedly prove that vowels "sung" in whisper in "the first five tones of a diatonic scale" display a substitute for rise in pitch: added noise in the sound spectra of [e, i, o] and a change in the location of formants in [a, u].[9] Meyer-Eppler reaffirms the validity of the statement made by Giet in *Lingua*, namely that a change in formants (i.e. the character of the vowel) may cause the impression of a change in pitch. This agrees with the well-known phenomenon that in ordinary speech, change in pitch may lead to a change in perceived vowel quality.[10]

In 1957, Meyer-Eppler's experiment with whispered "song" was, in all essentials, duplicated by C. M. Wise and L. P.-H. Chong[11]. They, however, found nothing in the phonetic substance that might indicate differences in "pitch." On the basis of the somewhat tenuous assumption that "the difference between the tones

[5] "Experimentelle Untersuchungen zum Mechanismus von Stimme und Gehör in der lautsprachlichen Kommunikation," *Forschungsberichte des Wirtschafts- und Verkehrsministeriums Nordrhein-Westfalen*, No. 221, 1955, pp. 18-19.

[6] "Das Flüstern in seiner physio-pathologischen und linguistischen Bedeutung," *Lingua*, IV (1954-55), p. 371.

[7] Franz Giet, "Kann man in einer Tonsprache flüstern?" *Lingua*, V (1955-56), pp. 372-381.

[8] See reference in footnote 3.

[9] "Realization of Prosodic Features in Whispered Speech," *Journal of the Acoustical Society of America*, XXIX (1957), pp. 104-106.

[10] Cf. E. A. Meyer, "Zur Tonbewegung des Vokals im gesprochenen und gesungenen Einzelwort," *Phonetische Studien*, 1896-97, p. 20.

[11] C. M. Wise and Lily Pao-Hu Chong, "Intelligibility of Whispering in a Tone Language," *Journal of Speech and Hearing Disorders*, XXII (1957), pp. 335-338.

of a pair of otherwise identical-sounding words in Mandarin is a difference of pitch" (no mention being made of the loudness feature), they concluded that "such words actually could not be distinguished from each other if spoken in a whisper." They accounted for the intelligibility of Chinese in whispered conversation by the role of the context, and used the same explanation for the fact that 62 % of 3824 utterances, with a possible choice among four items for each utterance, were recognized correctly. A choice among four distinctive word tones is probably relatively rare in Chinese, and until we know more about the test material which the authors used, it is difficult to know what 62 % represents as a recognition score. It is encouraging to see from the article in question that the same authors are conducting an experiment with words spoken in isolation. (The attention of the present writer was drawn to the article of Wise and Chong after he had himself carried out experiments with whispered word tones in various languages.)

In my opinion, experimental phonetics cannot be used to demonstrate what is and what is not perceived by the ear; in other words, experimental phonetics is incapable of bringing out all the physical correlates to language stimuli. I am therefore inclined to draw no conclusion from the instrumental experiments of Meyer-Eppler or Wise and Chong as far as perceptibility is concerned. I prefer direct perception tests, which have been regrettably rare in the study of word tones in whisper.

It does not seem very probable that substitutes for pitch stimuli are used in whispered speech, since the employment of such substitutes would demand of speakers a different technique from that which is used normally. A different technique, such as using more stress in lieu of higher pitch, would have to be acquired by observation and imitation. This would be very difficult, since the child learning to speak in whisper can have no clear idea of what the relation is between stress and intended pitch in that particular kind of speech. Pitch substitutes other than stress, e.g. changes in the character of the vowel, are still less directly observable. But it is useful to distinguish between substitutes for pitch and the "phonologization", in the absence of voicing, of features which are present even in voiced speech. It may even be that this is what Meyer-Eppler would call substitutes for pitch characteristics.

In the following, we report some recognition tests for word tones

in whispered speech which were carried out during the author's stay as a guest in the Department of Linguistics of Harvard University, in the spring of 1958. The experiments are listed according to the language involved.

1. Norwegian

a. Subject A. S. (female, Oslo dialect, linguistically trained), was given a minimal pair of words: /sva:re/ with Tone 1, 'the answer', and /sva:re/ with Tone 2, 'to answer' (in Norwegian orthography, *svaret* and *svare*). A coin was thrown; if it showed heads, the subject was to whisper the word with Tone 1; if tails, with Tone 2. This was done 33 times and tape-recorded, while a written record was kept of the throws. The subject then listened back to the tape-recorded 33 words, and indicated for each word the tone she thought she heard. Like all the other subjects, she had been requested to make a choice of tones even if she was in doubt. The whole series of 33 words was played back to the subject three times, and therefore 99 indications of perceived tone were given. Of these 65 % were "correct," i.e. in agreement with the Tone 1 vs. Tone 2 instructions produced by the coin.

b. Same subject, same procedure, but with 50 whispered words, to which she listened back twice, so that 100 indications of perceived tone were given.[12] Here 70 % were in agreement with the record kept of heads and tails in coin-throwing, as explained for experiment 1a.

c. Subject, the author (same dialect as A.S.). Same test as 1b. Recognition score: 61 %.

d. Subject R. M. (female, Stavanger dialect; not linguistically trained, but taught to denote tones as numbers 1 and 2). Same test as 1b. Recognition score: 71 %.

e. Subjects A. S. and M. K. J. (the author). A. S. was given a thoroughly shuffled deck of one hundred cards. The deck contained five instances of each of the following forms, which occur

[12] In order to make it impossible for the subject to remember what he or she had said in the beginning or at the end, the two first and the eight last words out of 60 were cut out. This left us with 50 words for counting the score.

in pairs that are minimal except for tone (the *-t* is purely orthographic):

sivet 'the bulrush'	*sive* '(to) ooze'
svevet 'the floating'	*sveve* '(to) float'
været 'the weather'	*være* '(to) be'
svaret 'the answer'	*svare* '(to) answer'
lånet 'the loan'	*låne* '(to) borrow'
bordet 'the table'	*borde* '(to) board'
huset 'the house'	*huse* '(to) house'
lyset 'the light'	*lyse* '(to) shine'
løvet 'the leafage'	*løve* 'lion'
høyet 'the hay'	*høye* 'high', adj.

The whispering was not tape-recorded, but listened to directly by M. K. J. turning his back on the subject who whispered, so that he should not be helped by gestures such as raising or lowering of eyebrows, bending or stretching of neck, etc. M. K. J. wrote down the words he heard. M. K. J.'s recognition score: 53 %.

f. Same procedure as in 1e, but this time M. K. J. whispered and A. S. listened. A. S.'s recognition score: 73 %.

Summary of Recognition Scores for Norwegian

Subject whispering	Subject listening and recognizing	Experiment	Score
A. S.	A. S.	a	65 %
A. S.	A. S.	b	70 %
M. K. J.	M. K. J.	c	61 %
R. M.	R. M.	d	71 %
A. S.	M. K. J.	e	53 %
M. K. J.	A. S.	f	73 %

As a 50 % score means pure randomness, it will be seen that the above recognition scores are not very convincing. However, they show some degree of distinction. The two poorest scores were obtained by the author (experiments c,e), who, however, had been the whispering subject in the experiment which came out with the highest score (f, 73 %).

2. Swedish.

Subject P. E. T. (male, southern Swedish dialect, linguistically trained). Same procedure as in 1b. Test words: *tagen* ('the repeated pulling', plural) and *tagen* ('taken', past participle), identical except for tone. Recognition score: 100 %.

The subject said he was never in doubt about the "tone" heard. He and the author agreed that what distinguished the pronunciation of the two words was a very marked stress on the second syllable in the case of the past participle. More extensive testing is of course needed before anything can be said about whispered tones in Swedish.

3. Slovenian.

a. Subject R. Z. (male, Ljubljana dialect, linguistically trained). Procedure as in 1b. Test words: *pet* ('five', numeral), tone: "long falling," and *pet* ('to sing'), tone: "long rising." Recognition score: 73 %.

b. Same procedure as 3a, except that the test words were *las* ('strand of hair', nom. sg.), tone: "long falling," and *las* ('of the hair', gen. pl.), tone: "long rising." Recognition score: 77 %.

c. Subject A. G. (male, Maribor dialect, not linguistically trained, but recognizing the tones in voiced speech as numbers 1 and 2). Test words: *rak* ('crab', dat. sg.), tone: 'long falling,", and *rak* ('crabs', nom. pl.), tone: "long rising." Same procedure as in 1b. Recognition score: 85 %.

d. Subject A. G. listening to the tape made in experiment 3a (R. Z. whispering). 100 items for recognition. Score: 74 %.

e. Subject A. G. listening to the tape made in experiment 3b (R. Z. whispering). 100 items for recognition. Score: 71 %.

The subjects pointed out that they had different dialects. They also complained of noise during the playback process, which may have caused lower recognition scores than would have been achieved in the case of better reproduction.

Summary of Recognition Scores for Slovenian

Subject whispering	Subject listening and recognizing	Experiment	Score
R. Z.	R. Z.	a	73 %
R. Z.	R. Z.	b	77 %
A. G.	A. G.	c	85 %
R. Z.	A. G.	d	74 %
R. Z.	A. G.	e	71 %

On inspection, the results seem to be more successful in the Slovenian case than in the Norwegian one. For a possible

explanation, compare the comments following the Table of recognition scores for Chinese, below.

4. Chinese (Mandarin).

a. Subject Mr. K. C. C. (anthropologist). Test material: *lha* ('to pull'), tone 1, and *la* ('to cut'), tone 2. The subject was told before the testing took place that in this experiment there would only be a choice between tones 1 and 2. Same procedure as in experiment 1b.[13] Recognition score: 83 %.

b. Subject Mrs. K. C. C. (anthropologist). Same procedure and test material as in experiment 4a. Recognition score: 79 %.

c. Same procedure as in experiment 1c. The test material consisted 120 whispered words, 12 occurring 10 times each. Randomness in sequence was secured by shuffling cards (see experiment 1e). Of the four word tones in Mandarin, only two were chosen for each minimal set. There were six minimal sets, each representing one tonal contrast: 1/2, 1/3, 1/4, 2/3, 2/4, 3/4. The subjects were informed about the special contrast applied to each of the minimal sets. The 12 set members were:[14]

Minimal set no.	Word	Translation	Tone no.
1	*lha*	to pull	1
	la	to cut	2
2	*ba*	eight	1
	baa	numeral adjunct for a noun	3
3	*la*	he (pron.)	1
	lah	a couch	4
4	*far*	to punish	2
	faa	method	3
5	*ma*	hemp	2
	mah	to scold	4
6	*yaa*	hoarse	3
	yah	to press down	4

[13] I was not able to discover any difference between this Chinese type of whisper and any other kind of whisper which I have heard, e. g. in English or in Norwegian, as far as voicelessness is concerned.—The nature of the Mandarin tones is as follows: Tone 1, high level; Tone 2, high rise; Tone 3, low fall-rise; Tone 4, high-to-low fall.

[14] I am indebted to Mrs. R. C. Pian, Harvard University, for supplying the words to fill the requirements stated above.

Mr. K. C. C. was whispering, and Mrs. K. C. C. was listening. Recognition score: 88 %.

The 12 % of "incorrect" answers were the following fourteen out of 120:

far	2	was	heard	as *faa*	3	six times
ɬah	4	—	—	— *ta*	1	three times
lha	1	—	—	— *la*	2	twice
mah	4	—	—	— *ma*	2	once
ta	1	—	—	— *ɬah*	4	once
ma	2	—	—	— *mah*	4	once.

d. Same procedure and test material as in experiment 4c, but this time Mrs. K. C. C. whispered and Mr. K. C. C. listened. Recognition score: 73 %.

The 27 % of "incorrect" answers were the following thirty-two out of 120:

ɬah	4	was	heard	as *ta*	1	nine times
far	2	—	—	— *faa*	3	seven times
lha	1	—	—	— *la*	2	six times
ba	1	—	—	— *baa*	3	five times
faa	3	—	—	— *far*	2	twice
ma	2	—	—	— *mah*	4	once
yaa	3	—	—	— *yah*	4	once
ba	1	—	—	— *baa*	3	once.

Both subjects thought they might have made a few mistakes in whispering, "Mrs. C. perhaps more than Mr. C." No tape recorder was used in experiments 4c and d. The room was quiet. Being a married couple, the subjects were well acquainted with each other's speech.

Summary of Recognition Scores for Chinese (Mandarin)

Subject whispering	Subject listening aud recognizing	Experiment	Score
Mr. K. C. C.	Mr. K. C. C.	a	83 %
Mrs. K. C. C.	Mrs. K. C. C.	b	79 %
Mr. K. C. C.	Mrs. K. C. C.	c	88 %
Mrs. K. C. C.	Mr. K. C. C.	d	73 %

It will be seen that the Chinese subjects obtained far better scores than the Norwegian and Slovenian ones. Besides the reasons already mentioned, attention should be drawn to the non-pitch concomitants of Mandarin tones. As described by

M. Swadesh,[15] the tones also manifest differences in stress and length, which may act as cues for their recognition in whispered speech.

The conclusion seems unavoidable that word "tones" are reflected somehow in whispered speech, and better in some languages (Swedish, Mandarin) than in others (Norwegian, Slovenian). *How* they are reflected is an exceedingly difficult question, since we do not seem as yet to have the appropriate means for psycho-physical investigation of whisper. This is not to say that the sound spectrograph should not be used for such purposes: it may reveal something of interest, but only after careful evaluation of a great number of recordings.

It should not be forgotten that word "tone" is not a purely chromatic phenomenon; the frequency modulation is probably always coupled with changes in stress (either physiological or physical intensity, or loudness, or any combination of these). Whether or not in some cases the degree of stress is distinctive and the pitch merely concomitant to stress, is a matter of interpretation. To speak of stress features as substitutes for pitch is misleading, as these features also exist in voiced speech, and the spectrograms of whispered speech which we have seen so far do not seem to indicate anything about the manner in which stress features take the role of pitch in whisper. It is true that a vowel tends to have added noise when whispered in an intended rising tone, but some languages—among them Norwegian—combine stress with a low tone and have weaker stress on higher tones. Still, one would guess that a Norwegian whispering Meyer-Eppler's "sung" vowels would add noise to higher tones. As for the change in the location of formants in [a, u], it should be remembered that even in normal song the character of the vowels tends to change with the pitch. On the other hand, if we consider whistling, we know that the tone is raised or lowered by changing the configuration and volume of the mouth cavity. This must have been what Meyer-Eppler had in mind when he told his subjects to maintain "the quality of a given vowel as well as possible" (see the study referred to in footnote 9).

We still know very little about the relationship between physiological (muscular) intensity and pitch produced, or between

[15] Morris Swadesh, "A Condensed Account of Mandarin Phonetics," *Travaux du Cercle linguistique de Prague*, VIII (1939), 213-216, p. 214.

physical intensity (amplitude) and pitch perceived. As we progress in the study of the combination of stress and pitch characteristics, we must be prepared for the discovery of rules which are not common to all languages.

University of Bergen.