ACOUSTIC AND PERCEPTUAL CORRELATES OF VIETNAMESE FOLK POETRY RHYTHMIC STRUCTURE

Anh-Thư T. Nguyễn
Mountain Creek, Sunshine coast, Australia
<thunguyen1972@gmail.com>

Abstract
This paper reports a study on the acoustic realization and the perception of the rhythmic structure of Vietnamese folk poetry. Ten speakers of Sài Gòn dialect recite four folk poems that were made up of three-word, five-word, six-word, seven-word, and eight-word lines. The acoustic analysis showed that the duration and intensity results mirror each other in indicating a strong iambic pattern of prominence, supporting the literature that a line of folk verse with even number of syllables tend to have a series of iambs and when there is an odd number of syllables, the line usually ends with an iamb, not an anapaest (Durand and Nguyễn, 1985). The perception results showed that listeners relied on duration cues in judging the rhythmic patterns of the poetic lines while intensity was not used. Also, majority of listeners were not finely tuned to these acoustic cues and only a few listeners could detect them in parsing the poetic lines into detail bi-syllabic iambic units.

Key words: folk poetry, rhythmic structure, acoustics, perception.

ISO 639-3 language codes: vie

1. Introduction
This paper reports a study on the acoustic realization and the perception of the rhythmic structure of Vietnamese folk poetry. Rhythm can generally be characterized as the repetition of patterned sequences of elements, often varying in prominence (Fraisse 1974, 1982). This characterization applies most strictly to certain forms of music and poetry; however, normal speech also exhibits tendencies toward rhythmic patterning — for example, in the alternation of stressed and unstressed syllables. In addition, listeners may perceptually exaggerate the rhythmicity of utterances. For example, English-speaking listeners hear interstress intervals as more evenly spaced than they actually are (Darwin & Donovan 1980; Donovan & Darwin 1979; Lehiste 1977). Although the claim that speech is characterised by a rhythmical component is universally accepted, there is far less consensus as to what should be identified as the rhythmic unit and as the carrier of rhythmic beat in spoken language.

To identify rhythm in speech, we have to search for something that is either structured or recurrent within a limited time range. This is the same as that stands at the base of the conception of metre in poetry. Metres fulfil the requirements in that they are simple structures composed of regular successions of stresses and syllables, resulting in alternations of stressed and unstressed syllables. Furthermore, stresses and syllables fall precisely within the relevant time range. In effect, the study of speech rhythm has long been associated with metrics and poetry. Aristotle, in The Art of the Rhetoric made an early attempt to describe the rhythm of language, using metric concepts to describe the different speech styles of the people (e.g. the iambic metre was said to be the rhythm of the common people, the trochaic that of rhetoricians, etc.).

Although the question of metre is still of interest today, many phoneticians of the last century have shifted the focus of research on two other issues. One consisted in finding an acoustic correlate of the perception of the rhythm beat: results in this field are not utterly uniform, but there seems to be general
agreement on the importance of the vowel onset (see Allen 1972 and 1975). The second issue concerns the
alleged regular occurrence of syllables and/or stresses, that retained most of the attention through the years.

The two categories of stress-timed and syllable-timed languages have been introduced by Pike (1945)
referring to the impression that stresses seem to occur at regular temporal intervals in English, while
syllables seem to have similar durations in Spanish. Abercrombie (1967) drew on the distinction, also on the
basis that the different rhythmic structure of these languages seems to be reflected by the metrical units
adopted in poetry: Germanic languages count the length of verses in feet, while Romance languages use the
syllable as the basic metric unit. This view had been of great influence in the following years, but various
instrumental experiments failed to give evidence of isochrony at the foot or at the syllable level.

After these failures, some linguists (Bertinetto 1977 and Dauer 1983, among others) attributed the
impression of stress-timing or syllable-timing to structural properties of languages, such as the absence vs.
presence of vocalic reduction and a complex syllabic structure. Relying on these theories, Ramus, Nesp &
Mehler (1999) and Grabe & Low (2002) proposed acoustic correlates of these phonological properties based
on vocalic and consonantal durations. The authors claimed that their measures allowed for a scalar
characterisation of languages on the basis of rhythm properties. The validity and the stability of the acoustic
correlates (soon re-baptised rhythm metrics) were soon tested in other studies, which gradually introduced
new variables, such as different speech rates (e.g. Dellwo & Wagner 2003). Some authors have proposed
modifications of the formulae (e.g. Dellwo 2006, and Benton 2010), or have applied the formulae to different
durations, such as voiced and unvoiced intervals (e.g. Galves et al. 2002) or feet and syllables (e.g. Wagner
&Dellwo 2004, and Asu & Nolan 2006), or have proposed new metrics that are based on different rationales (e.g. Bertinetto & Bertini 2008). Despite some criticisms and a few failures (e.g. Barry & Russo 2003, and Arvaniti 2009), these measures and the perspectives they offer have raised (and are still raising) a
growing interest within the scientific community: various authors have used them with the aim or in the
attempt of categorising different languages, different language varieties (e.g. Giordano & D’Anna 2010) and
even to detect the interference of the rhythm properties of L1 on productions by L2 speakers (e.g. White &
Mattys 2007). The latest developments in this field include some attempts to merge the two aspects of speech
rhythm (namely, the segmental and the accentual levels) into multi-layer models (Bertinetto & Bertini 2010,
and O’Dell et al. 2010).

This study examined the acoustic realization and the perception of the rhythmic structure of
Vietnamese folk poetry on the basis of Lehiste’s assumption (1984, 1986, 1990a) that “the suprasegmental
system of a language is crystallized in the metric structure of its traditional poetry”. Poetic meter is also
considered to be a strong indicator for the metrical pattern of the language in which the poetry is written
which are probably often violated in spontaneous speech. Thus, poetic speech can be called to be
characterized by a maximal level of rhythmical harmony” (Wagner, forthcoming). Therefore, it was found by
Lehiste (1990a) that within a poetic line, all metric feet have approximately the same duration in English and
Latvian poetry even though no evidence of isochrony at the foot or at the syllable level has been found in the
languages by instrumental experiments. A number of early experiments measuring feet in English have
shown that foot duration is proportional to the number of syllables they contain (Shen & Peterson 1962;
Bolinger 1965; Uldall 1971; see Lehiste 1977). This applied also to reiterant speech, used by Nakatani,
O’Connor & Aston (1981) to test isochrony.

In many languages, a poetic line is structured in terms of number and type of syllables. For example,
the epic folksongs of Serbia, the so-called deseterac, have ten syllable lines. The classic Balto-Finnic folk
song meter, the Kalevala meter, employs an eight-syllable line. In other languages, rhythm may be based on
a succession of stressed syllables. Germanic languages developed a metric of that kind. In stress-timed
poetry in contemporary English, metric feet consist of a stressed syllable and a flexible number of unstressed
syllables, whose duration is subordinated to the requirement that within a poetic line, all metric feet have
approximately the same duration (Lehiste 1990a). Lehiste (1990a) also found that Latvian tended to equalize
the duration of the metric feet while in Finnish the duration is determined by the quality of the syllables
constituting the foot.

Vietnamese folk poetry meter is conditioned by the length of lines and arrangement of tones within the
line. Vietnamese is a tonal language consisting of six tones which are designated even or flat (bằng) and
uneven or sharp (trắc). Uneven tones are also referred to as "oblique". The two even/flat tones are: high level (ngang) and low falling (huyền). The four sharp/oblique tones are: high rising (sắc), curve (hỏi), broken (ngã) and low falling (nặng). The number of tones, as well as their realisation, varies according to dialect: the Northern dialect has 6 tones while Central and Southern dialects have 1 tone less due to the historical merging of the curve (hoi) and broken (ngã) tones. The rule of distribution of tones in poetry — the aim of which is to give music to a poem — specifies the slots in a line to be filled with even/flat tones as opposed to those to be filled with oblique/sharp tones (Công Huyền Tôn Nữ 2001).

Vietnamese folk poetry (ca dao), particularly the two common folk meters: six-eight (lục bát) and “double-seven-six-eight” (song thất lục bát) was impressionistically noted to “reproduce the common cadence of Vietnamese speech, a rising rhythm – the iamb (a group of two syllables with the stress on the second syllable), and the anapaest (a group of three syllables, with the primary stress on the last syllable and sometimes a secondary accent on the first” (Durand and Nguyễn 1985). It is also observed that “in literary and also in modern spoken Vietnamese, there has been a strong tendency toward using a two-two syllable rhythm. Vietnamese syllables are usually grouped in two-two rhythmic groups. This tendency has been crystallized into poetic norms of four-syllable lines” (Liêm 1970, page 1). Nevertheless, no empirical acoustical studies have been done to verify the stress and rhythmic pattern of the language or its folk poetry.

Vietnamese has no system of culminative word stress; nevertheless, it is widely accepted that there is stress in the sense of accentual prominence at the phrasal level (Thompson 1965; Nguyễn Đăng Liêm 1970). Duration, intensity, full tonal realisation of accented syllables have been observed to be important parameters for describing stress in Vietnamese (Dỗ 1986; Chaudhary 1983; Hoàng & Hoàng 1975; Gsell 1980). Regarding the stress patterning in utterances, it is generally agreed by some researchers that there is an alternating pattern of strong and weak syllables. Thompson (1965) stated that the majority of the syllables have medium stress. In a sequence of syllables, alternating ones are slightly louder (but not in a distinctive manner): “each pause group has at least one heavy stress and weak stresses are fairly frequent in rapid passages, rarer in carefully speech” (p. 50). Jones and Huỳnh (1960) stated that “normally the stresses in a Vietnamese utterance are conditioned by the junctures,” and regarded the fundamental stress pattern of Vietnamese as consisting of the alternating occurrence of a strong and weak stress, with the last word of the phrase receiving a strong stress. Consistent with Jones and Huỳnh’s observation, it is remarked by Cao (2003) that due to the demarcative function of stress/accent in Vietnamese, native listeners tended to hear a juncture after a stressed syllable even though there is no such pause in reality as examined by spectrograms.

In recent studies on more carefully phonetically controlled and specialized sets of Vietnamese disyllabic compounds and reduplications, Nguyen and Ingram (2007a, b) found that there was at least a phonetic tendency for the right hand element of a disyllabic compound word to be more prosodically prominent by a number of relevant phonetic measures: greater tonal f0 range, higher intensity, greater duration of the second syllable, and formant measurements indicative of more centralized vowel nuclei (vowel reduction) on the first syllable. Nguyen (2010) investigated the rhythmic patterns in Vietnamese polysyllabic words by examining the rhythmic patterns and their acoustic correlates in polysyllabic reduplicative words (2-,3-,4-,5-,6- syllable pseudo-words). The results showed that there is a tendency of syllable coupling indicated mainly by syllable duration pattern and supported by the native listeners’ perception results, suggesting that polysyllabic words in Vietnamese tend to be parsed into bi-syllabic iambic feet with a rightward or retrograde rhythmic pattern.

2. Cadao - Vietnamese folk poetry- metric structure

Ca dao takes a wide range of forms, from two-syllable to six- and eight-syllable verse (lục bát). The range of ca dao includes children’s game songs, love songs, lullabies, riddles, work songs, and reveries about spiritual and social orders (Balaban 1980). The two most common folk meters are six-eight (lục bát) and “double-seven-six-eight” (song thất lục bát). It is estimated that 90% of folk poetry are in six-eight meter (Nguyễn 1992). Both verse forms can be traced to two basic meters: a rising rhythm, the iamb (a group of two syllables with the stress on the second syllable), and the anapaest (a group of three syllables, with the primary stress on the last syllable and sometimes a secondary accent on the first) (Durand and Nguyễn 1985). Vietnamese sayings and proverbs are often folk poems in miniature; they combine two iambs, two anapaests, one iamb, and one anapaest, or varying numbers of both. A line of folk verse tends to be made up
of an even number of syllables, a series of iambs in particular, when there is an odd number of syllables, the line usually ends with an iamb, not an anapaest.

The constituent unit of six-eight (lục bát) verse is a couplet in which the first line has six syllables and the second line eight syllables. Usually, both lines consist of iambics, and the stress falls on each even-numbered syllable: in such a case, they can be described as iambic trimeter and iambic tetrameter, respectively. On occasion, the first line departs from the norm and comprises two anapaests, instead: two equal, balanced hemistichs divided by a medial caesura (three | three). The last (or sixth) syllable of the first line rhymes with the last-but-two (or sixth) syllable of the second line (or, as an archaism, with the fourth syllable). What stamps six-eight verse with individuality, however, is its pattern of euphonic requirements, an ingeniously plotted alternation of “flat” (bằng) and “sharp” (trắc) tones. The scheme, as observable in a typical couplet, is diagrammed in figure 1: the double-headed arrow indicates rhyming, an open space represents an unstressed syllable, and F and S stand for flats and sharps. It will be noted that the tonal rules apply only to accented (even-numbered) syllables; that the same type of tones is assigned to syllables taking up the same position in both lines (flats for I-2 and II-2, sharps for I-4 and II-4, and flats for I-6 and II-6); and that flats predominate: they flank a single sharp in each line and monopolize the rhyme-fellows (I-6 and II-6) as well as the last two stressed syllables of the second line (II-6 and II-8).

![Six-eight verse diagram](adapted from Durand and Nguyễn 1985).

A six-eight couplet can stand by itself as a poem, and it often does in folklore. Six-eight verse amounts to a concatenation of such couplets. The needed linkage between couplets is that the last (or eight) syllable in the second line of one couplet rhymes with the last (or sixth) syllable in the first line of the next couplet. Each line rhymes with the next, but a new, different rhyme appears in every other line, in every couplet.

The constituent unit of double-seven-six-eight verse is a four-line stanza in which a double-seven doublet is followed by a six-eight couplet. Each line of the double seven doublet has seven syllables: an anapaest followed by two iambics (three | two | two). The disposition of rhyming syllables and prescribed tones is shown in figure 2 below. The double-seven doublet resembles the six-eight couplet in that the last syllable of the first line rhymes with the penultimate stressed syllable of the second line. In everything else, the double-seven is the opposite of the six-eight. Each line has an odd number of syllables. The rhyme-fellow carry sharp tones. Odd-numbered syllables are accented and embody a tonal scheme which, significantly, contradicts that of six-eight verse and no longer favours flat tones: three flats are balanced by three sharps. Figure 2 represents the double-seven-six-eight quatrain – the doublet and the couplet are joined by the rhyme-fellows II-7 and III-6. It should be noted that II-7 and III-6 carry flats; as a result, such tones still predominate within the rhyming scheme of the double-seven quatrain, although sharps more than hold their own in the first two lines: they equal flats in number and appropriate the rhyme syllables (I-7, II-5).
3. Production experiment

3.1. Linguistic materials
In order to pursue the aim of the present study, which examines the rhythmic patterns of three-word, five-word, six-word, seven-word, and eight-word lines, four folk poems were used. The investigated poems are in the appendix. The first is a popular six-eight poem titled “Tát nước đầu đình” (Fetching the water before the village communal house), which has 16 lines. Generally, the folk poem conforms to the prescribed metrical structure described in figure 1, except for the three syllables which varied from the prescribed tones (the prescribed tones should be even/flat ones but they are in uneven/sharp tones, được, in the third line, mượn in the eight line, and lúc in the tenth line). The second is the first four double-seven-six-eight stanzas of the “Chinh phụ ngâm” (The song of a soldier’s wife) by Đặng Trần Côn, rendered in Chữ Nôm (Vietnamese scripts) and in double-seven-six-eight folk meter by Phan Huy Ích. The four four-line stanzas consisted of 16 lines in total. Generally, the tones of the poem conform to the prescribed metrical structure described in figure 2 above, except for two syllables which have flat/even tones instead of sharp/uneven tones (thành in the fifth line and bình in the ninth line). The third item is a five-word folk poem which has ten lines. The fourth is four three-word riddles and sayings of about two, three and four lines each, which have 11 lines in total.

3.2. Subjects
Ten speakers of the Sài Gòn dialect (5 males, 5 females) who came from Hồ Chí Minh city participated in the study. They were either visitors or newly arrived immigrants to Australia and had been in Australia from 2 weeks to 4 years. Their age ranged from 38 to 70 years. Their education levels ranged from high school to higher degrees.

3.3. Procedure
Subjects were given the poems in print to practice reading before the recording. They were asked to read the poems in a natural manner. In fact, since these are popular poems, most of them have either read, heard or even learnt about them in schools. Recordings were made in a quiet room using sound recording and editing computer software PRAAT (Boersma and Weenink 2007) at 22050 Hz sampling rate.

3.4. Measurement
The acoustic parameters measured included syllable duration, line duration, duration of pause following each poetic line, intensity, vowel formant and vowel duration. Temporal variations of F0 and tonal shapes are obvious components but will not be treated here due to the nature of the linguistic material. The test items were segmented via the Emu Speech Tools, (Cassidy 1999). First, the Emu Labeller was used to mark the
edges of the target syllables and vowels, relying primarily on the spectrographic display in the Labeller. Then the Emu-R statistical software was used to extract the target duration (ms). Peak intensity (dB) in syllables and vowel first and second formants (Hz) and vowel duration were measured manually via Praat (Boersma and Weenink 2007).

3.5. Analysis
There were in total 530 lines of poem (16 lines of six-eight poem + 16 lines of double-seven-six-eight poem + 10 lines of five-word poem + 11 lines of three-word riddles and sayings read by 10 speakers).

A mixed (fixed and random) effects analysis of variance (ANOVA) model, using the restricted maximum likelihood method to estimate variance components was used to statistically analyse the data for each poem. The dependent variables were syllable duration, line duration, pause duration, tentative foot duration (grouping of two syllables), and intensity. The fixed effects were syllable or foot positions (in a line), lines (number of words in a line type), and speakers. The random effect was syllables. Tukey post-hoc tests were carried out to determine the significant differences among levels of the main factor of syllable or foot positions and lines when necessary.

3.6. Results

3.6.1. Duration
The three-way mixed effect ANOVA (syllable positions x lines x speakers) on syllable duration of the six-eight poem showed a significant effect for main factors syllable positions: F (7, 1120) = 80.03, p < .0001, lines: F(1, 1120) = 8.93, p <0.001, speakers F(9, 1120) = 44.52, p < .0001, and the interaction syllable positions x lines: F( 5, 1120) = 7.09, p < .0001, while other interaction effects were not significant (positions x speakers: F(63, 1120) = 1.26, p = 0.08 ns., speakers x lines: F(9, 1120) = 0.8, p = 0.5 ns., syllable positions x speakers x lines: F(45, 1120) = 0.9, p = 0.5 ns.)

Figure 3. Mean syllable duration (ms) of the six-eight poem. S1, S2, S3, S4, S5, S6, S7, and S8 are syllable positions in a line. The symbol * indicates significance at p < .001.

A post-hoc Tukey test for the interaction syllable positions x lines (figure 3 above) showed that the even syllables were significantly longer than its preceding odd syllables for both six and eight lines of the poem, suggesting a tendency of syllable coupling headed by the even syllables. The last syllables of the lines were longest (significantly longer than all other syllables in the lines), possibly indicating both a final lengthening effect and a right–headed accented effect if the same result is found for intensity.

The three-way mixed effect ANOVA results for the double-seven-six-eight poem showed a significant effect for main factors syllable positions: F (7, 1120) = 17.37, p < .0001 and speakers F(9, 1120) = 26.98, p <
.0001, but non-significance for lines: F(1, 1120) = 2.14, p = 0.1 ns., and the interaction syllable positions x lines: F(12, 1120) = 0.37, p=0.8 ns. Other interaction effects were not significant (positions x speakers: F(63, 1120) = 1.64, p = 0.5 ns., speakers x lines: F(18, 1120) = 1.67, p = 0.3 ns., syllable positions x speakers x lines: F(100, 1120) = 0.98, p = 0.7 ns.)

The post-hoc results of the interaction syllable positions x lines (figure 4 below) showed that the timing pattern of the six- and eight-word couplets mirrors that of the six-eight poem: even syllables are significantly longer than their preceding odd counterparts. In the seven-word lines, the odd syllables (S3, S5, S7) were significantly longer than their preceding even syllables, supporting the literature claim that the odd-numbered syllables are accented and the line has an anapaest followed by two iambics (three | two | two). The last syllable of the lines also had the longest duration as found in the six-eight poem. Also, there seems to be some type of line-medial prominence in double-seven-six-eight (figure 4), but not in six-eight poem (figure 3); particularly, in the eight-word lines, the fourth syllable was as long as the last syllable.

Figure 4. Mean syllable duration (ms) of the double-seven-six-eight poem.
S1, S2, S3, S4, S5, S6, S7, and S8 are syllable positions in a line.
The symbol * indicates significance at p < .001.

The two-way mixed effect ANOVA (syllable positions x speakers) on syllable duration of the five-word poem showed a significant effect for main factors syllable positions: F (4, 500) = 7.64, p < .0001, speakers F(9, 500) = 23.55, p < .0001, but the interaction syllable positions x speakers was not significant: F(36, 500) = 1.15, p = 0.2 ns.

A Tukey post-hoc test showed that the second syllable was significantly longer than the first syllable. The third syllable was significantly longer than the first and the second syllable. The fifth syllable was significantly longer than the fourth syllable and also the longest, while the fourth syllable was the shortest. This result suggests a rhythmic pattern of an anapaest followed by an iamb.
The two-way mixed effect ANOVA (syllable positions x speakers) on syllable duration of the three-word riddles and sayings showed a significant effect for main factors syllable positions: F (2, 330) = 13.09, p < .0001, speakers F(9, 330) = 16.50, p < .0001, but the interaction syllable positions x speakers was not significant: F(18, 330) = 1.27, p = 0.05 ns.

A post-hoc test showed that the first syllable was significantly longer than the second syllable. The third syllable was the longest, significantly longer than the first and the second syllable. This can suggest either an anapaest pattern or a tentative two-foot pattern in which the first syllable stand by itself and was lengthened to fill the foot template and the other two syllables constitute an iamb. This needs to be verified in a perception study to see how listeners parse the rhythmic pattern of the line.

3.6.2. Intensity
The three-way fixed effect ANOVA (syllable positions x lines x speakers) on syllable intensity of the six-eight poem showed a significant effect for main factors syllable positions: F (7, 1120) = 10.26, p < .0001, lines: F(1, 1120) = 35.09, p < .0001, speakers F(9, 1120) = 195.93, p < .0001, and the interaction syllable
positions x lines: $F(5, 1120) = 2.81, p < .02$, while other interaction effects were not significant (positions x speakers: $F(63, 1120) = 1.44, p = 0.01$ ns., speakers x lines: $F(9, 1120) = 1.54, p = 0.1$ ns., syllable positions x speakers x lines: $F(45, 1120) = 0.92, p = 0.6$ ns.)

**Figure 7.** Mean syllable intensity (dB) of the six-eight poem. S1, S2, S3, S4, S5, S6, S7, and S8 are syllable positions in a line. The symbol * indicates significance at $p < .001$.

A post-hoc Tukey test for the interaction syllable positions x lines (Figure 7 above) showed that the even syllables had significantly higher intensity than their preceding odd syllables for both sixth and eighth lines of the poem. This matches the duration results, suggesting a pattern of syllable coupling headed by the even syllables. The last syllables of the lines also had very strong intensity, together with the duration results, suggesting a right-headed accented effect.

The three-way mixed effect ANOVA results for the double-seven-six-eight poem showed a significant effect for main factors syllable positions: $F(7, 1120) = 10.50, p < .0001$, lines: $F(1, 1120) = 5.85, p < .05$, speakers $F(9, 1120) = 101.45, p < .0001$, and the interaction syllable positions x lines: $F(12, 1120) = 3.17, p < .001$, while other interaction effects were not significant (positions x speakers: $F(63, 1120) = 1.29, p = 0.2$ ns., speakers x lines: $F(18, 1120) = 1.17, p = 0.3$ ns., syllable positions x speakers x lines: $F(100, 1120) = 0.81, p = 0.8$ ns.)

The post-hoc results of the interaction syllable positions x lines (figure 8 below) showed that for the six-word and eight-word lines, the even syllables had higher intensity than their preceding odd counterparts. Nevertheless, the significant patterns were less robust than that of the duration results. The reason the fourth syllable of the eight-word line had lower intensity is that these syllables in the double-seven-six-eight poem carry either the low falling tone (nặng) or the curve tone (hỏi) which tend to be creaky and thus had low intensity values. There was not a significant intensity pattern for the seven-word lines because many accented syllables (S5 and S7) carry the low-falling (nặng) and curve (hỏi) tones, which have low intensity due to creakiness.
The two-way mixed effect ANOVA (syllable positions x speakers) on syllable intensity of the five-word poem showed a significant effect for main factors syllable positions: F(4, 500) = 5.60, p < .001, speakers F(9, 500) = 187.46, p < .0001, but the interaction syllable positions x speakers was not significant: F(36, 500) = 1.63, p = 0.24 ns.

A Tukey post-hoc test showed that the third syllables had higher mean intensity than the first and the second syllables, however it did not reach significance. The fifth syllable was significantly higher in intensity than the fourth syllable and the fourth syllable had the lowest intensity. This, together with the duration result, suggests a rhythmic pattern of an anapaest followed by an iamb.

The two-way mixed effect ANOVA (syllable positions x speakers) on syllable intensity of the three-word riddles and sayings showed a significant effect for main factors syllable positions: F (2, 330) = 6.1, p < .01, speakers F(9, 330) = 9.87, p < .0001, but the interaction syllable positions x speakers was not significant: F(18, 330) = 0.16, p = 1 ns.

A post-hoc test showed that the first syllable was significantly higher in intensity than the second syllable. The third syllable significantly had higher intensity values than the second syllable. Particularly, the
first syllable had the strongest intensity. This, together with the duration result, suggests a tentative two-foot pattern in which the first syllable was lengthened and strongly accented to fill the foot template and the other two syllables constitute an iamb. However, this needs to be verified in a perception study to see how listeners parse the rhythmic pattern of the line.

Figure 10. Mean syllable intensity (dB) of the three-word riddles and sayings. S1, S2, and S3 are syllable positions in a line. The symbol * indicates significance at p < .001.

In summary, the duration and intensity results mirror each other in showing a strong iambic pattern of prominence, supporting the literature claiming that a line of folk verse with even number of syllables tend to have a series of iambs, and that when there is an odd number of syllables, the line usually ends with an iamb, not an anapaest (Durand and Nguyễn 1985). This suggests that syllables in Vietnamese folk poetry tend to be grouped into a tentative iambic bi-syllabic foot and when there is only one odd syllable, the syllable tends to be lengthened to fill the bi-syllabic foot template. If native listeners’ parsing pattern matched this acoustic prominence pattern, it would suggest that a bi-syllabic foot is a prosodic unit above the syllable in Vietnamese folk poetry meter, which is the aim of the perception experiment reported in a later section of this paper.

3.6.3. Vowel reduction

In order to examine whether there is phonetic reduction associated with the examined acoustic pattern of asymmetric prominences, a small follow-up experiment was carried out. Vowels of the first syllable (unaccented syllable) and second syllable (accented syllable) of the tentative iambic foot in the six-eight poem were compared with matching vowels elicited in a control condition in which speakers said the target word in a sentence “Đọc lại từ ----- đi nhé” (Please say the word ----- again). Two of the ten speakers (one male and one female) were asked to record the control vowels which were then compared with their counterpart vowels in the poem. The vowels investigated included: /i/ /ɛ/, /e/, /a/, /ʊ/ and /o/ embedded in the targets words as shown in Table 1 (below). The target tentative foot used for this vowel analysis were bolded in the six-eight poem in the appendix of stimulus material. First and second formants of the vowels under investigation were taken at vowel mid-point manually via Praat. Vowel duration was also measured. The result is shown in figures 11 and 12 (below).
Figure 11. Vowel plots produced by the male (top fig.) and female (bottom fig.) speakers. Vowel SAMPA symbols without a number: control condition. Vowels with number 1: target word in the first unaccented position of a two-word foot in the 6-word or 8-word poetic line. Vowels numbered 2: target word in the second accented position of a two-word foot in the 6-word or 8-word line. SAMPA symbols: i: /i/, E: /e/, e:/ɛ/, a:/a/, 6:/u/, o:/o/, and O:/ɔ/.
Table 1. Vowels and target words

<table>
<thead>
<tr>
<th>Vowels</th>
<th>Control words</th>
<th>1st unaccented words</th>
<th>2nd accented words</th>
</tr>
</thead>
<tbody>
<tr>
<td>/i/</td>
<td>Thi /tʰi/21/</td>
<td>Thi /tʰi/21/ (thì cho)</td>
<td>chi /ci/313/ (sứt chí)</td>
</tr>
<tr>
<td>/e/</td>
<td>Quên /wen/33/</td>
<td>Trèn/ tʂen/33/ (trên cânh)</td>
<td>quên /wen/33/ (bỏ quên)</td>
</tr>
<tr>
<td>/ɛ/</td>
<td>me /me/ 322/</td>
<td>mɛ /me/322/ (mẹ già)</td>
<td>-</td>
</tr>
<tr>
<td>/a/</td>
<td>qua /wa/33/</td>
<td>trạ /tʂa/313/ (trả công)</td>
<td>qua /wa/33/ (hôm qua)</td>
</tr>
<tr>
<td>/o/</td>
<td>cânh /kɐɲ/21/</td>
<td>ān /un/33/ (ān rau)</td>
<td>cânh /kɐɲ/21/ (trên cânh)</td>
</tr>
<tr>
<td>/ɔ/</td>
<td>hôm /hom/33/</td>
<td>hôm /hom/33/ (hôm qua)</td>
<td>công koŋm/33/ (trả công)</td>
</tr>
<tr>
<td>/ɔ/</td>
<td>cho /cɔ/33/</td>
<td>cho /cɔ/33/ (cho cùng)</td>
<td>cho /cɔ/33/ (thì cho)</td>
</tr>
</tbody>
</table>

As shown in figure 11, there was not much distinction between the vowels of the unaccented first syllable and those of the accented second syllable of the foot. For the male speaker, both vowels were more centralised in comparison with the control vowels. For the female speaker, both vowels in the orally produced poem tended to be more back in comparison with their counterparts in the control condition. This result only indicated that vowels in poetry readings tend to be somewhat altered or reduced in acoustic quality compared with when they were produced in a more carefully speaking style, a phenomenon predicted to be more pervasive in spontaneous Vietnamese than already documented and needs to be more systematically studied.

A mixed effect two-way ANOVA (conditions x speakers) was conducted on vowel duration. The fixed effect was conditions (control vs. 1st unaccented syllable vs. 2nd accented syllable) and speakers. The random effect was vowels. The results showed significant effects for conditions (F(2,41)= 5.4, p<0.02), speakers (F(1, 41) = 23.7, p<0.01), and the interaction conditions x speakers (F(2, 41)=4.2, p<0.05). A post-hoc test showed that the 1st unaccented vowel was significantly shorter than the control vowel and the 2nd accented vowel for both speakers as indicated in figure 12. This result indicates that while the 2nd syllable in the tentative bi-syllabic foot was lengthened when accented, the first unaccented syllable was reduced, probably to fit in a foot timing template as investigated in the section on foot duration below.

Figures 12. Mean vowel duration (ms) . The symbol * means p < 0.01.

3.6.4. Foot duration
In order to examine whether the tentative iambic feet are isochronous, a three-way ANOVA (foot positions x lines x speakers) was conducted on the duration of bi-syllabic iambic feet (duration of a bi-syllabic foot = duration of 1st syllable + duration of 2nd syllable) of the six-eight poem. The result showed a significant effect
for main factors foot positions: $F(3, 560) = 31.84, p < .0001$, lines: $F(1, 560) = 5.95, p < .02$, speakers $F(9, 560) = 25.06, p < .0001$, and the interaction foot positions x lines: $F(2, 560) = 9.5, p < .0001$, while other interaction effects were not significant (positions x speakers: $F(27, 560) = 0.79, p = 0.76$ ns., speakers x lines: $F(9, 560) = 0.48, p = 0.8$ ns., foot positions x speakers x lines: $F(18, 560) = 0.9, p = 0.56$ ns.)

A post-hoc Tukey test (figure 13) showed that only the last foot of the six- and eight-word lines was significantly longer than all the other preceding feet in the line respectively, while the non-final feet were not significantly different in duration from one another, indicating a pattern of foot isochrony. The non-final feet of the six-word line were also found to be equal (not significantly statistically different) to non-final counterparts of the eight-word line. This strongly suggests an isochronous bi-syllabic foot-timing template.

![Figure 13. Mean foot duration (ms) of the six-eight poem; f1, f2, f3, and f4 are foot positions in a line. The symbol * indicates significance at p < .001; ns. means non-significance.](image)

The three-way ANOVA (foot positions x lines x speakers) on the duration of tri-syllabic and bi-syllabic iambic foot of the double-seven-six-eight poem showed a significant effect for main factors foot positions: $F(3, 501) = 66.3, p < .0001$, lines: $F(2, 501) = 50.41, p < .0001$, speakers $F(9, 501) = 19.48, p < .0001$, and the interaction foot positions x lines: $F(4, 501) = 85.2, p < .0001$, while other interaction effects were either weakly or not significant (positions x speakers: $F(27, 501) = 2.3, p < .05$, speakers x lines: $F(18, 501) = 2.6, p < .05$, foot positions x speakers x lines: $F(36, 501) = 1.1, p = 0.28$ ns.)

A post-hoc Tukey test (figure 14) showed a similar pattern of foot isochrony for the six-eight couplets as that of the six-eight poem above. That is, only the last foot of the six-word line was significantly longer than all the other preceding feet in the line respectively, while the non-final feet of both lines were not significantly different in duration from one another, strongly confirming a pattern of foot isochrony regardless of the segmental structure of the disyllabic sequence, consistent with Latvian and English metric feet (Lehiste 1990a). On the other hand, in the seven-word line, the first three-syllable foot was much longer than its following bi-syllabic feet, suggesting foot duration increases with the number of syllables, consistent with Shen & Peterson (1962), Bolinger (1965), Uldall (1971) and Nakatani et al (1981). It is wondered how listeners parse the three-syllable anapaest of the seven-word lines: as one unit of three words together or as two units one | two.
Figure 14. Mean foot duration (ms) of the double-seven-six-eight poem; f1, f2, f3, and f4 are foot positions in a line. The symbol * indicates significance at p < .001; ns. means non-significance.

3.6.5. Line and pause duration
In order to examine the higher-level timing pattern above the line level, two-way ANOVAs (lines x speakers) were conducted on the line and pause duration of the double-seven-six-eight poem. The results showed significant main effect of lines (6-word vs. 7-word vs. 8-word): line duration: $F(3, 150) = 41$, $p < .0001$, pause duration: $F(3, 150) = 10.14$, $p < .0001$ and speakers: $F(3, 150) = 7.5$, $p < .0001$ and $F(3, 150) = 2.83$, $p < .0001$ respectively. The interaction of lines x speakers were not significant: $F(27, 150) = 1.9$, $p =0.26$ ns. for line duration data and $F(27, 150) = 0.8$, $p =0.5$ ns. for pause duration data).

A post-hoc test (figure 15) showed that the six-word line was the shortest, the seven-word line was significantly longer and the eight-word line was the longest. In the seven-word doublet, the second seven-word line was significantly longer than the first line. This indicates that with isochrony at the foot level in Vietnamese six-eight poem, there was no isochrony at the line level as found in Lehiste (1990b)’s study in Lithuanian poems. In terms of pause duration, the pauses following the second seven-word line and the eight-word-line were significantly longer than those following their preceding counterpart: the first seven-word line and the six-word line respectively. Nevertheless, the pause following the eight-word line (i.e. that last line of the stanza) was not significantly larger than that of the second seven-word line. This result indicates that speakers tended to organise the double-seven-six-eight poems into units of doublets and couplets instead of four-line stanza units.

In summary, the duration and intensity results mirror each other in showing a strong iambic pattern of prominence, supporting the literature claiming that a line of folk verse with even number of syllables tend to have a series of iambs and when there is an odd number of syllables, the line usually ends with an iamb, not an anapaest (Durand and Nguyễn 1985). The results of this experiment also mirror the findings on acoustic and perceptual rhythmic pattern of polysyllabic word (Nguyen 2010) and supporting the iambic acoustic pattern in a subset of bi-syllabic compounds and reduplications investigated (Nguyen and Ingram 2007a, 2007b). This suggests that syllables in Vietnamese folk poetry tend to be grouped into an iambic bi-syllabic foot. This is strongly supported by the isochronous iambic foot pattern regardless of the segmental structure of the disyllabic sequence in the six-eight metric. It is likely that “poetic speech maximizes the language specific rhythmical constraints which are probably often violated in spontaneous speech. Thus, poetic speech can be said to be characterized by a maximal level of rhythmical harmony” (Wagner, forthcoming). The next section reports the results of the perception test, which aimed to examine whether the acoustic cues are employed in the perception of the rhythmic structure of the folk poetry by native listeners.
4. Perception experiment

4.1. Stimuli
Stimuli of two speakers (1 male and 1 female) from the production experiment were used for the perception study. The criteria used for the selection of stimuli were that the speakers had an easy-to-hear voice, which was judged to be clear, loud enough to be heard by the experimenter. There were 106 utterances to be judged ((16 six-eight lines + 16 double-seven-six-eight lines + 10 five-word lines + 11 three-word lines) x 2 speakers). The tokens were put in random order each with two immediate repetitions with a gap of about 6 seconds between each item.

4.2. Subjects
Twenty subjects of the Southern dialect (8 males, 12 females) with no known auditory deficiencies participated in the perception experiment. Listeners in this study were all tertiary students between 23 to 39 years of age.

4.3. Procedure
In the perception test, subjects listened to the test tokens twice each and were asked to judge the rhythmic pattern of the poetic line and put a slash after the syllable where they thought the rhythmic beat fell. A sample of the answer sheet is as follows:

Hôm qua / tát nước / đầu đình /

The experiment was carried out using a laptop computer with loud speakers in a quiet classroom at Can Tho University located in the delta to the southwest of Sài Gòn.

4.4. Analysis
There were five data sets of poetic lines (three-word lines, five-word lines, six-word lines, seven-word lines, and eight-word lines). Separate mixed effect ANOVAs were conducted for each data set. The dependent variable was the percentage number of responses (the number of listeners over the total listeners) for each stimulus token. The independent variables were rhythmic patterns and speakers. The random effect was poetic lines or stimulus tokens.

A second type of analysis involved the correlation between the acoustic parameters, namely duration and intensity, and the perception scores. Perception scores which were equal to the proportion of listeners who marked that syllable/word as having a stress beat were calculated for each syllable in the poetic lines,
namely, three-word, five-word, six-word, seven-word and eight-word lines. Then Pearson correlations were computed between the acoustic parameters and the perception scores for each set of the poetic line data. Results are presented in Table 2 at the end of the next result section.

4.5. Results

4.5.1 Three-word lines

The ANOVA result on three-word poetic line set showed no significant effect for the main factors: rhythmic patterns ($F(1,44)=0.33$, $p=0.62$, ns) and speakers ($F(1,44)=0$, $p=0.98$, ns), but the interaction effect patterns x speakers was significant: ($F(1,44)=12$, $p<0.01$).

As shown in figure 16, two rhythmic patterns were judged for the three-word poetic lines: the rhythmic beat fall on the first syllable or on the third syllable. Nevertheless, the agreement rate among listeners for each rhythmic pattern is not high: there was no significant difference between the two rhythmic patterns for the male speaker’s stimuli, there was a significant effect for female speaker’s stimuli but the difference is not large (56% on first syllables vs. 44% on third syllables). This result shows that the three-word lines can be heard as either two rhythmic units (one | two) by some listeners or one unit of three words together by other listeners, indicating that the intensity and duration cues found on the first syllable, though salient acoustically, can only be detectable in perception by some Vietnamese listeners but not all. On the other hand, this may also indicate that the three-word line is not long enough to carry at least one repetition of the basic unit of rhythm. It is found by neurophysiological and psychoacoustic research that there exist “sliding windows of temporal integration” of approximately the size of 150-200ms and a window of roughly 400-600ms can be regarded as temporal present and the time span of the temporal present co-occurs with typical foot length across a variety of languages (Wagner, forthcoming). The mean duration of the three-word lines is 862.53ms, which is greater than the “temporal present.”

![Figure 16. Mean response percentage of the three-word line data set. 1 and 3 are rhythmic patterns (1: the rhythmic beat judged to fall on the first syllable or 3: on the third syllable). The y-axis: mean response percentage. * indicates significance at p < 0.01.](image)

4.5.2 Five-word lines

The ANOVA result on five-word poetic line set showed a significant effect for the main factor rhythmic patterns ($F(3,57)=19.5$, $p<.0001$) but no significant effect for speakers ($F(1,57)=0.02$, $p=0.9$, ns) and the interaction effect patterns x speakers: ($F(3, 57)=0.59$, $p=0.7$, ns). As shown in figure 17 below, four different rhythmic patterns were perceived for the five-word poetic line: on the second syllable (16%), on the third syllable (50%), on the fifth syllable (19%) and on the first, third, and fifth syllables (14%). Even though
the agreement rate was not very high, generally a majority of the five-word lines were parsed on the third syllable: a three | two pattern, consistent with the duration and intensity results and supporting the literature of an anapaest followed by an iamb pattern. Only the last four lines of the poem were parsed on the second syllable with a two | three pattern by some listeners. This could be due to the syntactic structure of the lines. Apparently few listeners could attend to the fine acoustic cues in parsing the lines at 1st 3rd and 5th syllables or into one | two | two pattern.

![Five-word poem](image)

**Figure 17.** Mean response percentage of the five-word line data set. 2, 3, 5 and 135 are rhythmic patterns. The y-axis: mean response percentage. * indicates significance at p <0.01.

4.5.3 Six-word lines

The ANOVA result on six-word poetic line set showed a significant effect for the main factor rhythmic patterns (F(3, 62) = 7.8, p < .0001) and the interaction effect patterns x speakers: (F(2, 62) = 8.5, p < .0001) but no significant effect for speakers (F(1,62) = 0.7, p = 0.3, ns). As shown in figure 18 below, only 46% listeners could detect the acoustic cues (duration and intensity) in parsing the six-word poetic lines in detail into two-syllable iambs (at 2nd, 4th, and 6th syllables or two | two | two pattern). By contrast, 42% listeners parsed the lines at the 2nd syllable and 12% parsed at the 4th syllable.

![Six-word poem](image)

**Figure 18.** Mean response percentage of the six-word line data set. 2, 4 and 246 are rhythmic patterns. The y-axis: mean response percentage. * indicates significance at p <0.01.
4.5.4 Seven-word lines
The ANOVA result on seven-word poetic line set showed a significant effect for the main factor rhythmic
patterns (F(4,42) = 95, p < .0001) but no significant effect for speakers (F(1,42) = 0.17, p = 0.7, ns) and the
interaction effect patterns x speakers: (F(3, 42)= 0.11, p =0.9, ns.). As shown in figure 19, 81% of listeners
parsed the seven-word lines at the 3rd syllable and into a three | four pattern, while only 7% of listeners
parsed at the 3rd and 5th syllables into a three | two | two pattern, or as an anapaest followed by two iambs, as
suggested by the acoustic results and as claimed by the literature (Durand and Nguyễn 1985).

![Figure 19. Mean response percentage of the seven-word line data set. 3, 13, 35 and 1357 are rhythmic
patterns. The y-axis: mean response percentage. * indicates significance at p <0.01.](image)

4.5.5 Eight-word lines
The ANOVA result on eight-word poetic line set showed a significant effect for the main factor rhythmic
patterns (F(3,69) = 47, p < .0001) but no significant effect for speakers (F(1,69) = 0.27, p = 0.8, ns) and the
interaction effect patterns x speakers: (F(3, 69)= 1.3, p =0.3, ns.). As shown in figure 20, 62% of listeners
parsed the eight-word lines at the 4th syllable consistent with the line-medial prominence in the acoustic
results, while only 15% of listeners could detect the acoustic cues in parsing the lines at 2nd, 4th, 6th, and 8th
syllables or into two-syllable iambs as suggested by the acoustic result.

![Figure 20. Mean response percentage of the eight-word line data set. 3, 13, 35 and 1357 are rhythmic
patterns. The y-axis: mean response percentage. * means significance at p <0.01.](image)
4.5.6 Correlation results

As shown in Table 2 below, there was a significant correlation between duration and perception scores for all poetic lines while the correlation between intensity and perception scores was not significant. This indicates that listeners relied on duration cues in judging the rhythmic patterns of the poetic lines while intensity was not used.

Table 2. Results of Pearson correlation between duration and intensity and perception scores, ns: non-significance.

<table>
<thead>
<tr>
<th>Poetic lines</th>
<th>Duration</th>
<th>Intensity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Three-word lines</td>
<td>r = 0.45, p &lt; 0.001</td>
<td>r = 0.16, p = 0.19 ns.</td>
</tr>
<tr>
<td>Five-word lines</td>
<td>r = 0.37, p &lt; 0.001</td>
<td>r = 0.06, p = 0.51 ns.</td>
</tr>
<tr>
<td>Six-word lines</td>
<td>r = 0.27, p &lt; 0.001</td>
<td>r = 0.11, p = 0.16 ns.</td>
</tr>
<tr>
<td>Seven-word lines</td>
<td>r = 0.33, p &lt; 0.001</td>
<td>r = 0.13, p = 0.17 ns.</td>
</tr>
<tr>
<td>Eight-word lines</td>
<td>r = 0.28, p &lt; 0.001</td>
<td>r = 0.05, p = 0.49 ns.</td>
</tr>
</tbody>
</table>

4.6. Discussion and conclusion

The perception results showed that listeners relied on duration cues in judging the rhythmic patterns of the poetic lines while intensity was not used. Even though the acoustic results suggest iambic patterns of prominence for the poetic lines, supporting the literature claim about the poetic rhythmic pattern (Durand and Nguyễn 1985) and strongly implying that syllables in Vietnamese folk poetry tend to be grouped into iambic bi-syllabic foot, the perception experiment showed that majority of listeners were not finely tuned to these acoustic cues and only a few listeners could detect them in parsing the poetic lines into detail bi-syllabic iambic units.

Why listeners in Nguyễn (2010) could parse the polysyllabic reduplicative pseudo-word into iambic patterns as reflected in the duration and intensity cues but most listeners of this study could not parse the poetic lines in detail is probably due to the structure of the stimuli. Stimuli in Nguyễn (2010) were polysyllabic words with constant segmental makeup and tones, while the syllables in poetic lines in this study were of different segmental composition and tones. This is consistent with the result of a perception study (Nguyễn and Ingram, submitted) in which Vietnamese listeners listened to disyllabic words (disyllabic compounds, phrases, full reduplications and tone sandhi reduplications) and judged which syllable of the sequence (the first or the second) was more prominent or if both syllables were of equal prominence. The results show that subjects performed at chance level for most of the disyllabic word types and only the tone sandhi forms (with constant segmental makeup but with an alternation of tones) were judged to have more prominent on the second syllable. This contradicts results showing that listeners in other languages, such as English, can hear stress contrasts regardless of syllable segmental composition (Fry 1958). This suggests that prosodic asymmetry in Vietnamese is a ‘sub-perceptual’ threshold phonetic effect and can only be detected by some individuals with phonetically fine-tuned ears.

In the production study, the duration and intensity results mirrored each other in showing a strong iambic pattern of prominence, supporting the literature claiming that a line of folk verse with even number of syllables tend to have a series of iambs and when there is an odd number of syllables, the line usually ends with an iamb, not an anapaest (Durand and Nguyễn 1985). The results of this experiment also mirror the findings on acoustic and perceptual rhythmic pattern of polysyllabic word (Nguyen 2010) and support the iambic acoustic pattern in a subset of bi-syllabic compounds and reduplications investigated (Nguyen and Ingram 2007a, 2007b). This suggests that syllables in Vietnamese folk poetry tend to be grouped into an iambic bi-syllabic foot. This is supported by the isochronous iambic foot pattern regardless of the segmental structure of the disyllabic sequence in the six-eight metric, consistent with Latvian and English metric feet (Lehiste 1990a). On the other hand, in the seven-word line, the first three-syllable foot was much longer than its following bi-syllabic feet, suggesting foot duration increases with the number of syllables, consistent with Shen & Peterson (1962), Bolinger (1965), Uldall (1971), and Nakatani et al. (1981). It is likely, that “poetic speech maximizes the language specific rhythmical constraints which are probably often violated in
spontaneous speech. Thus, poetic speech can be said to be characterized by a maximal level of rhythmical harmony” (Wagner, forthcoming).

Acknowledgements
I would like to thank the subjects for their voluntary participation in the experiments. Thanks to Mrs. Tho Nguyen for administering the perception experiment. Thanks to Dr. John Ingram and the anonymous reviewers for their insightful feedbacks.

References


Appendix: Folk poems

NB: The bolded words were those used for the analysis of vowel formants.

(1)  
**Tat nuoc dau dinh poem**

Hôm qua tát nước đầu đình

Bỏ quên cái áo trên cành hoa sen

Em được thì cho anh xin

Hay là em để làm tin trong nhà?

Áo anh sứt chỉ đường tà

Vỡ áo chưa có, mẹ già chưa khâu

Áo anh sứt chỉ đã lâu

Mai muốn có ấy về khâu cho cùng.

Khâu rồi anh sẽ trả công

Đến lúc có chồng anh lại giúp cho

Giúp em một thùng xôi vò

Một con lợn béo, một vò rượu tâm.

Giúp cho đôi chiếu em nằm

Đôi chăn em đắp, đôi tằm em đeo

Giúp em quan tâm tiền cheo

Quản năm tiền cuối lại đểo bước cau.

(2)  
**The first four stanzas of the Chinh Phu Ngam**

Thuở trời đất nổi cơn gió bụi

Khách má hồng nhiều nỗi truân chuyên

Xanh kia thăm thẳm tầng trên

Vì ai gây dựng cho nên nỗi này

Trống Trường Thành lung lay bóng nguyệt

Chín tầng gươm báu trao tay

Nửa đêm truyền hịch định ngày xuất chinh

Nước thanh bình ba trăm năm cũ

áo nhung trao quan vũ từ đây

Sứ trời sớm giục đường mây

Phép công là trọng, niềm tây sá nào

Đường giong ruổi lưng đeo cung tiễn

Buổi tiễn đưa lòng bận thê noa

Bóng cờ tiếng trống xa xa

Sầu lên ngọn ái, oán ra cửa phòng

(3)  
**Five-word folk poem**

Ai qua cầu rơi dép

Ai xuống cầu nhặt giúp

Nhặt ba lần không nản
Vấn kính cận ung dung  
Biết người có trí lớn  
Là một đấng anh hùng  
Trao cho sách Thái ất  
Về nhà học thuộc lòng  
Ngày sau ra giúp nước  
Lập được đại kỳ công

(4) Three-word riddles and sayings
Một mẹ nằm  
Trăm con bước.  
bằng cái vung  
Vùng xuống ao  
Đào không thấy  
Lấy không được.  
Đói ăn rau  
Đau uống thuốc.  
Vào mồng ba  
Ra mồng bẩy  
Rẫy mồng tám.