

**Kakinuma, Toshie: «Jōji Yuasa and the Apparatus Called Language», in: *Ongaku Geijutsu*, Issue 10, October 1984, pp. 100–103.**

Toshie Kakinuma (born in Shizuoka, Japan) is a Japanese musicologist, critic, and translator specializing in 20th- and 21st-century experimental and vocal music. She studied musicology at Kunitachi College of Music (B.M., 1977) and Ochanomizu University (M.A., 1981), before earning her Ph.D. in Music from the University of California, San Diego in 1989, where she conducted research on the American composer and instrument builder Harry Partch. After returning to Japan, she taught at several universities, including Meiji Gakuin and Takushoku, and later served as Professor of Musicology at Kyoto City University of Arts, where she also directed the Archival Research Center. Her research focuses on American experimentalism, contemporary Japanese music, organology, and the relationship between sound, language, and the body.

Kakinuma is the author of *American Experimental Music as Ethnic Music* (Film Art Sha, 2005) and numerous essays on composers such as Yūji Yuasa, Tōru Takemitsu, and Lou Harrison. She has also translated major works of 20th-century music literature into Japanese, including *John Cage's Silence* and *Alex Ross's The Rest Is Noise*.

At the time she wrote “Yūji Yuasa and the Apparatus Called Language” (originally published in *Ongaku Geijutsu*, October 1984), she was a young critic and researcher deeply engaged with the new directions of postwar Japanese vocal and electronic music.

Jōji Yuasa (1929–2024) was a Japanese composer and one of the pioneers of postwar Japanese avant-garde music. He studied composition in Japan and Europe and was influenced by Western modernist trends (like Olivier Messiaen, Pierre Boulez, Karl-Heinz Stockhausen). He was known for his experimental, combining Western avant-garde techniques with Japanese aesthetics. He worked extensively with serialism, electronic music, and open-form composition.

Yuasa's Work with the voice was an exploration of vocal timbre. He was interested in the human voice as a sound source, not just as a carrier of text or traditional melody. He often treated the voice like an instrument, exploring extended techniques (whispering, shouting, overtones, glissandi), integrating it in *Musique Concrète* and *Electronic Music*, experimenting with tape manipulation of the voice, cutting, layering, and processing vocal sounds to create textures and sonic landscapes. The voice became material for composition, abstracted from semantic content. He composed pieces for solo voice, mixed ensembles, and electronic setups. Vocal performances were often experimental, and performers had to explore unconventional sounds and interactions with electronics or tape. Like Tōru Takemitsu, Yuasa was interested in the “sound itself”, focusing on timbre, gesture, and expressive potential of the voice. The voice could evoke natural or bodily phenomena or serve as an abstract sonic object in avant-garde composition.

Jōji Yuasa is a central figure in postwar Japanese experimental music who treated the human voice as an instrument and sonic material, exploring both acoustic possibilities and electronic transformation. His work aligns with broader trends in Japanese musique concrète, emphasizing sound over meaning.

### Abstract Part I

Toshie Kakinuma's essay "Jōji Yuasa and the Apparatus Called Language" analyzes the Japanese composer Jōji Yuasa's 1969 tape piece *Voices Coming* as a radical experiment in the relationship between sound, language, and communication. Kakinuma situates the work within the postwar avant-garde tradition of deconstructing words and meaning – alongside Boulez, Berio, and Stockhausen – yet argues that Yuasa's approach is distinct in its phenomenological rigor. Rather than mixing speech with electronic or musical sounds, Yuasa confines himself to unaltered recorded voices, exposing the hidden musicality of ordinary speech. By isolating interjections, conjunctions, and conversational fillers from everyday dialogue, *Voices Coming* strips language of semantic function and re-presents it as temporal and formal sound material. Through this process, habitual speech becomes defamiliarized: language, usually transparent and functional, regains perceptible form and duration. Kakinuma interprets this transformation as a double act – emptying everyday meaning while enriching non-everyday significance – and as a key moment in Yuasa's broader search for a synthesis of poetic and musical temporality. The essay concludes that Yuasa's work transcends the conventional boundaries of 'music' and 'vocal composition', revealing how the human voice mediates between sound, meaning, and the possibility (or impossibility) of communication.

Jōji Yuasa's *Voices Coming*<sup>1</sup> has often been called a 'problematic work.' In particular, the second piece, *Interview [Intervyū]*, has received a wide range of evaluations: Some have doubted 'whether it can even be heard as music,' while others have argued 'it is rather a form of poetry.' Some critics regard it as an important work that defines a turning point in the composer's career, or even that with this work Yuasa "stood at the forefront [*saizenei*]" of musical innovation; yet it has also been dismissed by others as a 'failure.' Yet to ask whether *Voices Coming* is or is not 'music' seems, in the end, almost meaningless. For today's *ongaku* [*music*], has already departed from the reified, fetishized realm of traditional *ongaku* or *seigaku* [*vocal music*].

The history of vocal works since the 1950s has been one in which language was stripped of meaning – a history that, in Pierre Boulez's words, "sentenced poetry to the punishment of

---

<sup>1</sup> Jōji Yuasa: *Voices Coming* (1969) – Tape composition for recorded voices, ca. 25 minutes, realized at the NHK Electronic Music Studio, Tokyo. The work consists of three parts:

I. *Telephonopathy* – fragments of international telephone calls

II. *Interview* – everyday Japanese speech with interjections and fillers ('ano', 'sono', ect.).

III. *In Memory of Two Slain Peace Warriors* – overlapping speeches by Inejiro Asanuma (1898–1960 leader of the Japanese Socialist Party) and Martin Luther King Jr. (1929–1968), civil rights activist, use of non-violent resistance and civil disobedience against discrimination.

Together, the three sections trace a progression from communicative to purely sonic language.

dismemberment.”<sup>2</sup> Karlheinz Stockhausen, in his electronic piece *Gesang der Jünglinge* (1956), decomposed the boy’s recitation of the Old Testament into syllables [onsetsu] and phonemes [onso], synthetically reconstructing them so that words could be apprehended as sound-objects [onkyō sozai]. Berio’s *Thema – Omaggio a Joyce* (1958)<sup>3</sup> likewise fragmented readings of *Ulysses* in English, French, and Italian, mixing them with electronic sounds to present ‘unintelligible words.’

Boulez, in *Cummings ist der Dichter* (1970),<sup>4</sup> transferred the spatiality of e. e. cummings’s calligram poems into music, while Lutosławski, in *Trois poèmes d’Henri Michaux* (1963),<sup>5</sup> used invented words, retaining only the formal shell [fōrumu] of language, to create a world reminiscent of a Balinese *kecak* chant. In such works, the human voice is no longer merely a medium for words. Berio’s *Sequenza III* (1965) and *Visage* (1961)<sup>6</sup> pursued every possible vocal gesture – laughter, shouting, tongue-clicks, breath sounds – that could be produced orally.<sup>7</sup>

These acts of “shattering words [kotoba no funsai]” and “abolishing meaning [imi no haiki]” were linked to the assertion that the distinction between the two apparatuses [sōchi] – language and music – should be withdrawn. After this, it even became something of a trend in avant-garde [zen’ei] music.

– page 100 –

<sup>2</sup> Vgl. Boulez, Pierre: “Sonata, What Do You Want of Me?”, in: Pierre Boulez: *Notes of an Apprentice*, trans. Herbert Weinstock, New York: Alfred A. Knopf, 1968, pp. 41–61.

<sup>3</sup> Luciano Berio: *Thema (Omaggio a Joyce)*, for voice and tape; based on a recorded reading by Cathy Berberian of a section of *Ulysses* (Chapter *Sirens*) by James Joyce, [8:09], 1958.

For Berio, the human voice retains theatrical immediacy – this may be described as its ‘corporeal reality.’ For Yuasa, by contrast, the voice has become an ‘acoustic trace’ [onkyō no ato], a memory of language.

<sup>4</sup> Pierre Boulez: *Cummings ist der Dichter*, for 16 solo voices and chamber ensemble, [10’], 1970, based on E. E. Cummings *birds*, No. 64 from *No Thanks* (1935).

<sup>5</sup> Witold Lutosławski: *Trois poèmes d’Henri Michaux*, for mixed choir and orchestra of winds, brass, harp, pianos, percussion, [20’], 1963.

<sup>6</sup> Luciano Berio: *Visage*, for electronic sound and Cathy Berberian’s voice, [21’], 1961, see Part I footnote 2) Luciano Berio: *Sequenza III*, Solo female voice (unaccompanied), [9’], 1965, see Part I footnote 7)

<sup>7</sup> The result of *Voices Coming* is that the work no longer concerns communication between speaker and listener, but rather exposes the very structure of speech itself as a sonic phenomenon [onkyō-genshō]. In this sense, Yuasa’s treatment of the voice can be contrasted with Berio’s *Visage* (1961). Whereas Berio constructs a dramatic continuum through the montage of Cathy Berberian’s expressive vocal gestures, Yuasa removes all traces of dramatic persona or narrative context.

In this transformation lies the crucial difference between European *musique concrète vocale* and Yuasa’s approach, which Kakinuma describes as *genko no jiko hansha* – a ‘self-reflection of language

Yuasa's first work for voice dates back to *Aoi no Ue* (1961).<sup>8</sup> Although I have not had the opportunity to hear this piece, according to the composer it is based on a libretto that paraphrases a *Noh* text – “a text such that even if the parts that function as prose and their context are dissolved, it can still stand musically or poetically; in that sense, there are portions that transcend words themselves.”

That the composer, who studied [*yōkyoku*] chanting in his youth, should have begun his vocal works by engaging with such traditional material is a noteworthy fact.

After *Aoi no Ue* came *Voices Coming* [*Voicesu Kamōngu*], 1969. From this point onward, Yuasa's works for voice began to take an increasingly radical direction. There followed by a series of works for voice: *Question*<sup>9</sup> and *Utterance*<sup>10</sup> (both 1971), *Calling Together*<sup>11</sup> (1973), *Projection on Bashō's Haiku*<sup>12</sup> (1974), *Projection on Onomatopoeia*<sup>13</sup> (1979), and *Etude for “The”*: *Observation on Weather Forecast*<sup>14</sup> (1983) – a continuous line of inquiry into the relationship between voice and language.

Leaving aside [*Aoi no Ue*] for the moment, Yuasa's experiments with the voice had already taken a decisively radical turn with [*Voices Coming*].

He stated of *Voices Coming*: “Here I focused primarily on *spoken language* – *parole*<sup>15</sup> – and composed while considering the problem of communication through sound and language.

---

<sup>8</sup> Jōji Yuasa: *Aoi no Ue*, Electronic Music, tape, [27'], 1961, with the voices of three performers of the Kanze school: Kanze Hisao, Kanze Eio, and Kanze Shizuo.

In *Aoi no Ue*, the human voice [*ningen no koe*] becomes material [*sozai*]. Here lies, according to Kakinuma, Yuasa's concern: not in the meaning of words [*kotoba no imi*], but in their sound aspect [*onkyō-men*]. Words are dissolved into sound; in that sense, *Aoi no Ue* can be regarded as a predecessor to *Voices Coming* (1969). At the same time, the *Noh* text still provides a framework. The composer states that he used the libretto only as a ‘point of departure,’ not as a means of narrative expression. In *Voices Coming*, this textual support disappears entirely; it is composed solely from fragments of ordinary speech, isolated and reassembled through tape manipulation.

<sup>9</sup> Jōji Yuasa: *Question*, for mixed chorus, [16'], 1965.

<sup>10</sup> Jōji Yuasa: *Utterance*, for mixed chorus, [18:30], 1971

<sup>11</sup> Jōji Yuasa: *Yobikawashi* [*Calling Together*], for mixed voices, [10'-20'], 1973. A ‘theatre piece’ exploring linguistic concreteness and performative speech.

<sup>12</sup> Jōji Yuasa: *Projection on Bashō's Haiku*, for mixed chorus and vibraphone [14'], premiered by the Tokyo Mixed Chorus, November 1974.

<sup>13</sup> Jōji Yuasa: *Projection Onomatopoeic*, mixed chorus, [15'], 1979.

<sup>14</sup> Jōji Yuasa: *Observation on Weather Forecast*, bariton, tp, [8'], 1983.

<sup>15</sup> *parole* – term from Saussurean linguistics, used by Yuasa in French transliteration – in Saussure's sense: the individual act of speaking, as distinct from *langue*, the abstract system of language.

Between voice and word, between meaningful and meaningless utterance, there exist, so to speak, various intermediate conditions – various intermedia problems.”<sup>16</sup>

This statement by Yuasa about [*Voices Coming*] shows little difference from Berio’s remark that “there is something more musical than musical sound — the sound of the speaking voice and noise.”<sup>17</sup> Yet whereas Berio, when incorporating recorded conversation into his work, uses it as part of an amalgam with other prose texts and electronic sounds, Yuasa’s method is far more thoroughgoing. In *Voices Coming*, the composer restricts himself entirely to *raw recorded speech* as material.

*Voices Coming* (1969) consists of three parts. In the first part, *Telephonopathy* [*Telefonopashī*], we hear various calling words transmitted through international telephone circuits — “[*moshi-moshi*] ... *hello* ... *Tokyo* ...” and so on. The second piece, *Interview* [*Intabyū*], consists almost entirely of conjunctions and interjections that hold little or no meaning — phrases such as “[*anoo*] ... [*maa*] ... [*itte mireba*] ...” etc. The third piece, *In Memoriam of Two Assassinated Peace Warriors* [*Korosareta futari no heiwa senshi o kinenshite*], intertwines speeches by *Asanuma Inajirō* and *Martin Luther King Jr.*<sup>18</sup>

Expressions like “hello?” [*moshi-moshi*] possess, by virtue of their parallelism, a certain poetic character; and speech delivered as oratory is often said to be the most poetic form within everyday language. What the Russian Formalists discovered in Lenin’s language was precisely the poetic character of the speech form. However, the words in the second piece are taken directly from ordinary conversation. They are neither poetry nor prose, nor newly coined words. Why, then, did Yuasa focus on “spoken language”? What, for him, was the experience of spoken and everyday language?

– page 101 –

In everyday language [*nichijō gengo*], attention is ordinarily not directed toward the qualities of words or the *grain of the voice*<sup>19</sup>. What is of foremost importance is the semantic content,

<sup>16</sup>Jōji Yuasa: Program notes / LP liner notes; later in: Joji Yuasa: *Collected Works Vol. 2*, 1999.

<sup>17</sup> For more on Berio’s thoughts on the ‘boundaries between language and music,’ see: Luciano Berio: “Two Interviews,” in *Contemporary Composers on Contemporary Music*, ed. Schwartz & Childs, 1967.

Luciano Berio: “Meditations on a Twelve-Tone Horse,” in *Incontri Musicali*, no. 3, 1956.

<sup>18</sup> See footnote 1)

<sup>19</sup> For further philosophical implication of the wording *grain of the voice* see Roland Barthes (1915–1980). Barthes was a French literary theorist, philosopher and semiotician. In his 1972 essay *Le grain de la voix* (“The Grain of the Voice”), Barthes explores the material, tactile dimension of vocal sound beyond semantic content. He

that is, the matter to be conveyed, or else the very fact that an exchange of words is taking place. Of course, the delicate nuances and intonations of words function as powerful elements that can affect meaning, and they can act with great efficacy in actual linguistic behavior.

However, needless to say, in daily conversation words are not exchanged for the purpose of enjoying their inflections or accents.

*Interview [Intabyū]* exposes the hidden musicality of language, usually unnoticed by us, by cutting away the meaningful parts from everyday language, which ought primarily to convey meaning. It is, so to speak, the reverse side – the negative – of language. The subsidiary ornaments of language, which could not exist independently in the space of daily life, here come to possess autonomous significance. The everyday meaning is stripped away, and a new meaning is generated. In this way, there arises a twofold act – “the `emptying of everyday language and the enrichment of non-everyday meaning`.”

Then what, after all, is the autonomous meaning of the sounds or meaningless parts of words, or the meaning made non-everyday?

To state the conclusion first: it is the recognition of the *form* and *temporality* of language [*kotoba no fōrumu to jikanteki ninshiki*]. The world of *Voices Coming* is one that makes the form and temporality of words perceptible.

Everyday language – that is, the words that have become habitual to us – through repeated use loses its vividness and becomes transparent like air. For the sake of ease of perception, it slips through the ear and leaves no trace. In other words, familiar words do not make us feel form or time; rather, they are closely bound to place and space.

The method of *Voices Coming* transfers words from the series of space [*kūkan no keiretsu*] / `ordinary series` to the series of time [*jikan no keiretsu*] / `poetic-musical series`,<sup>20</sup> thereby acquiring a new freshness of perception. Words that would normally pass straight through the

---

distinguishes between the *message* of the voice and its *material* quality, the “grain” – that texture which carries the presence of the body and the moment of utterance. Barthes argues that the voice’s grain is what escapes meaning and carries a subtle affective potential.

<sup>20</sup> [*kūkan no keiretsu* / *jikan no keiretsu*] – literally “series of space” / “series of time,” Yuasa’s terminology for perceptual domains.

ear – “anoo... maa... itte mireba... sono!”<sup>21</sup> – are fully examined by the listener, and the experience is prolonged. Perception is extended and fully brought to consciousness. Words are apprehended through their form and acquire temporality. Moreover, the several utterances of “anoo” by person A produce a softened effect of parallelism [*pararerizumu no kōka*], which may then be compared with B’s “anoo” or C’s “sono!” Conjunctions or meaningless words that at the moment of conversation could not be apprehended in either form or time are thus placed within a *temporal series*.

This process differs in kind from the operations of many avant-garde composers who cut and reconstruct poems or prose that already exist within the flow of time.

If [*Yuasa*] can be said to stand at the ‘forefront,’ it is in this sense.

Let us listen, then, to *Interview* [*Intabyū*]<sup>22</sup> with closer attention.

– page 102 –

*Interview* is divided into six sections corresponding to six questions. In the first part a single male speaker uses many words beginning with a vowel – “ano..., *arui wa, itte mireba...*”

Section two features both male and female voices.

The voices of different people overlap; and one can discern changes in tone color.

The man frequently uses words beginning with a consonant such as “sono..., *sore wa..., tatoeba...*”

In section three the number of speakers increases to three or four, heightening both complexity and tension. In the fourth section, once again two speakers – a man and a woman – appear. Their voices, different from those in the second part, form a contrasting pair.

In the fifth section, the scene passes to a single male speaker whose utterances are sparse, producing a sense of silence and interruption within the flow.

In the sixth section, the same male speaker is questioned again, he replies, “I feel like I could answer, but the moment I do, it somehow feels like it's not quite right,” in a coda-style punchline [*kōda-fū no ochitsuki*].

---

<sup>21</sup> [*anoo*], [*maa*], [*itte mireba*], [*sono!*] – Japanese conversational fillers or hesitation forms.

<sup>22</sup> See footnote 1)

When one listens in this way, the *grain of each person's voice* [*koe no kime*], their manner of speaking, individual features, and inflections can be accurately grasped and calculated within the entire musical structure.

This process reveals that, in such a procedure, words regain their own distinctive and autonomous form and time.

But at that moment, the time of words is no longer merely the time of words.

It is simultaneously the manifestation of poetic time and musical time.

– page 103 –

(to be continued)

Toshie Kakinuma – music critic

Editorial note:

This text is a transcription of the original Japanese article by Toshie Kakinuma and does not constitute a full translation. It was transcribed, rendered, and edited by Dorothea Schürch using Google Drive OCR, DeepL, and ChatGPT. Source: Toshie Kakinuma, “Jōji Yuasa and the Apparatus Called Language” [湯浅譲二と言語という装置], *Ongaku Geijutsu* 音楽芸術, vol. 42, no. 11 (Nov 1984), pp. 100–103. Abstract and footnotes by Dorothea Schürch.