情報処理学会研究報告
IPSJ SIG Technical Report

Vol.2009-SE-164 No.14
Vol.2009-EMB-13 No.14
Vol.2009-CSEC-45 No.14
2009/5/29

SPIT 判別のための チューリング・テスト方式の研究

松 倉 俊 介† 佐 藤 直†

概要：FTTH などの普及により，IP 電話の利用者が増加しているが，一方で SPAM メールのような迷惑 IP 電話の出現が予測されている．この IP 電話版 SPAM は SPIT(SPAM over IP Telephony)と呼ばれている．本研究は，SPIT が発信から通話までをすべて自動（無人）で行われるものと仮定し，その自動判別手法を提案する．具体的には，質問応答による論理型判別を行うチューリング・テストを提案する．SPIT 側が自動応答システムを利用して応答することを想定し，SPIT 判別上望ましいと思われる質問構成法を考案した．さらに，現状で利用可能な自動応答システムを用いて応答特性を評価した．検討の結果，本提案による質問構成法が SPIT の自動判別に有効であるという見通しを得た．

A proposal of the Turing test against SPIT

Shunsuke Matsukura† Naoshi Sato†

Abstract According to the expansion of broadband access networks such as FTTH, the number of IP telephone service users is increasing. However, similar to SPAM mails, the users may suffer from embarrassing advertisement calls in the future. SPAM in use of IP telephone is called SPIT, standing for SPAM over IP Telephony. This study assumes that entire process including dialing and speech for SPIT would be automated by programmed machines, and proposes a Turing test method to classify the sender of IP call into a real person or the machine. In the proposed method, a question is given from the receiver to the sender, and response of the sender in answering the question is investigated. Especially the study discusses how to compose sentences for the question. Further it simulates the method with an auto-answer system and examines classification performance. The simulation result shows that the proposed method is prospective to counter SPIT.

† 情報セキュリティ大学院大学 情報セキュリティ研究科

† 情報セキュリティ大学院大学 情報セキュリティ研究科

1. はじめに

インターネットは目覚しい普及，発展を続けており，高度情報化社会のインフラとして，重要な位置を占めるに至っている．その発展により，通信手段として用いられていた電話に，IP 電話の誕生をもたらした．音声をデータ圧縮，符号化して IP パケットとして比較的に安価なコストメリットがあり，今後の普及が期待されている．

一方，普及が進むにつれ，直面する大きな課題が予想されている．既にインターネット上の通信手段として広く用いられている電子メールには，SPAM と呼ばれる，主に広告(製品やサービスの売り込み)メッセージが大量に送信されてくるという問題が発生している．企業などの団体ではこうしたSPAMを未然に防ぐ取組みはしっかり行われているが家庭や個人においてはまだまだ対策は十分とは言えず，多くの電子メール利用者がSPAMを受信せざるを得ない状況にある．

このSPAMがIP電話においても，今後発生すると予測されている．IP電話におけるSPAMは，SPIT(SPAM over IP Telephony)と呼称されている．今のところ，まだ被害件数はほとんどない．しかし今後，SPAM配信者が矛先をIP電話に向けてくる可能性は高い．

2. SPIT について

2.1 SPIT の概要

SPAM の一般的な定義は配信される事を望んでいないメッセージ，または，非常に大きな(容量的な意味で)メッセージを示す．IP 電話における SPAM とは，既存の加入電話における迷惑電話に当たる．現在の社会で，迷惑電話自体がそこまで深刻な問題になっていない．しかしながら，電子メール SPAM と同じく，SPIT は非常に問題視されている．その理由を，既存の加入電話，電子メール SPAM などと比較しつつ，以下に記述する．

本論文では，SPIT 自体は人間の操作を必要としない，電話の発信から通話までをすべて機械による自動制御のプログラムであると仮定し，そのような SPIT の対策を検討する．

2.2 SPIT の問題点について

既存の加入電話における広告電話は人手によるためコストがかかる事が弱点である．しかし，IP 電話を用いると，呼接続や送話が自動化しやすくなることから，ハードウェアコス

Vol.2009-SE-164 No.14
Vol.2009-EMB-13 No.14
Vol.2009-CSEC-45 No.14
2009/5/29

トや人的コストは比較的安価に抑え，無作為に大量に広告電話を発信する事ができると考えられている．こうして電子メール SPAM と同じく，広告電話が大量に発信された場合，広くの受信者が被害を受けると考えられる．
また，SPIT は電子メール SPAM と違い，リアルタイムな通信である事も問題である．メールと違い，電話が着信した場合，すぐに受話器を取るのが慣習となっているため，広告としての即時性が高いこと，また，メールの Subject やプレビューに相当する機能がないために，電話に出るまでその内容もわからないことから，迷惑電話の不快感は電子メール SPAM を受け取った時よりも大きく，時間の浪費感も強く感じる事になる．
さらに，電子メール SPAM と同じく，大量に発信される事により，ネットワークへの負荷も深刻なものになる．普通，電子メールは 1 件あたり 1〜10KB の容量であるが，SPIT を音声データによるボイスメールと仮定すると 1 件あたり約 100KB となる．電子メール SPAM と同等に広まれば，そのネットワーク負荷も現在の 10 倍以上になる．
このような問題が予測される SPIT ではあるが, 現在までに被害はほとんどない. これは, 技術的・経済的問題があり，まだ商用化レベルに到っていないからであると推察される．IP 電話がさらに普及し環境が整えば，SPIT は急速に広まるものと予測される．

2.3 先行研究

NEC が開発した VoIP SEAL が代表的なものとして挙げられる[1]．このシステムは，人間からの電話とスパム生成ソフトを用いた電話とを，SIP サーバ上でのチューリング・テストによって区別し，SPIT と判別されれば通話を切断するという内容となっている．
チューリング・テストとは，イギリスの数学者であるアラン=チューリングが考案し，人と機械を見分けるための方法を総称していうテストの方式を言う．VoIP SEAL で用いるチューリング・テストの内容は，発信者に対してシステムが録音されたメッセージを流し，そのメッセージに対して発信者が決まった応答パターンを返せるかをテストするものである．人間には電話の時に特定の会話パターンがあり，そのパターンを音声エネルギーの変化として，システム側で決まったパターンに従うかどうかを監視している．

2.4 電子メール SPAM 対策とその応用

電子メールにおける SPAM 対策は社会的にも SPAM に対する認知が広まった結果, 様々な対策がなされている．ここではその主な対策について記述し，それらの SPIT 対策への応用について検討する．

2.4.1 フィルタリング

電子メールの SPAM 対策として，まず挙げられるのがテキストベースのフィルタリング機能である[1]．フィルタリングの基準は技術によって様々で，一部のものは，正規表現によるマッチングに基づいて判断する．他のものとしては，ヘッダや本文内のキーワードを探す手法, 送信者やサーバのアドレス（IP アドレス含む）をチェックする手法などがある．また，テキスト以外のコンテンツによるブロックも存在する．例えば，画像フィルタの場合，いわゆるアダルト画像を複雑な画像解析によって識別するなどの利用法がある．

SPIT への応用を踏まえると, 電話がリアルタイム通信である事により, メールのように, 閲覧前に本文等を精査して分別するフィルタリングを行う事が出来ず，既存のメールにおける SPAM 対策とは異なった視点で対策を考えなければならない．
電話において，特定の発信に対して，着信要求を破棄したり，着信元のルールなどによる着信拒否を行う技術は当然存在する．問題は，ある発信に対してその内容を特定する事が難しい事にある．電話がリアルタイムな通信である事から，電話の受信者は，着信した電話を取るまではその中身を知る事は出来ない．要は，電話を取るまで SPIT かどうかを判別する事が出来ないという事が問題となる．

また，電話が着信した時点で，発信者が広告などの目的を果たせずとも，受信者にとっては SPIT に成り得る．

SPIT 対策としての重要な課題の 1 つは, ある電話の発信に対して, その発信が SPIT か, 正常な発信者かを判別できるであり，SPIT を機械と判別するチューリング・テストが必要になる．チューリング・テストが有効で, SPIT 判別の見通しが得られれば, 判別した SPIT に対して，以降の通信を拒否するなどにより，受信者を負担をかけない方法で処置できると考えられる．

以上から，本稿においては，機械と人を見分けるチューリング・テスト方式を提案する事とする．

2.4.2 CAPTCHA 技術

CAPTCHA は大量のメール送信を防ぐための措置である．電子メール SPAM では，送信元を偽装しているものも多いが，フリーメールアドレスを取得して大量に電子メール SPAM を送信するものもある．こうしたフリーメールアドレス取得を自動的にさせないた

Vol.2009-SE-164 No.14
Vol.2009-EMB-13 No.14
Vol.2009-CSEC-45 No.14
2009/5/29

め，CAPTCHA と呼ばれる画像認証技術が利用されている．

CAPTCHA の画像認証では，認証する側が多少読み取りにくい数字や文字を含んだ画像をユーザに示し，ユーザがそれらを目視で読み取り返信する．また，画像を目で読み取る事が出来ない障害者向けに音声を読み上げて入力を促す音声 CAPTCHA のような技術も存在する．これも機械と人を区別するチューリング・テストと言える．

電話が音声による通信である事から，画像 CAPTCHA をチューリング・テストとして使う事は出来ないので，音声 CAPTCHA を SPTI 対策に応用する事が望ましい．しかし，音声 CAPTCHA は 2008 年 5 月に 8 割ほどの確率で破る事が出来ると報告されている[3]. このため，音声 CAPTCHA をそのまま SPIT に対するチューリング・テストは突破される可能性が大きく，何らかの改善が必要となる．

2.5 論理型認証によるチューリング・テスト

音声 CAPTCHA は，前述のようにかなり高い確率で破られているが，その理由は文字や数字の単純な読み上げであった事が指摘されている．その読み上げパターンを解析すれば回答できることが原因と思われる．これを踏まえて，本研究では，より人間的な知的能力を必要とする論理型認証を組み合わせた音声応答方式を検討する．

論理型認証は質問文に対して，その回答を答える方式である．SPIT 判別のためのチューリング・テストとして用いる場合，電話の発信者に対し，質問文を流し，発信者はそれに対して応答する形とする．

以下では，チューリング・テストの内容として，質問文の内容と，回答方式を検討する．

2.6 提案方式のための要件

論理型認証で用いる質問文をチューリング・テストとして用いる場合，その難易度は，発信者が人である事も踏まえ，なるべく簡潔かつ，短いものでなければならない．
提案法のチューリング・テスト内容の要件として，本稿では以下を設定することとする．
①論理型認証の対象は電話の発信者とする．
②受信者がオフフックし会話が可能になる前の接続過程で実施する．
③質問は短く，発信者が人間ならば即在に回答可能で，
機械には回答が難しい，と見做されるものとする．
④質問に対して，送信者はプッシュボタンを用いて数字で回答できる．
⑤質問文としては，その内容以外から，回答候補を流す必要がない．

なお，上記の④と⑤は，発信者（回答者）が発声する必要が無く，システムとしても大規模とならないことを目的としている．

3. 提案法の概要

3.1 質問文の形式

発信者に対しての質問は，一般的な論理的思考力があれば即座に回答可能である程度のものとする．

また，当然ながら機械(SPIT)では容易に回答できないようなものとしなければならない．現在の機械の，質問に対する回答能力は質問応答技術の動向から推測可能と考えられる．また，SPIT 実行者は，こうした質問応答技術を汎用的なツールとして利用し，広告電話を大量送信する事が予測される．

以上のことから，本稿では，現在 WEB 上でテスト公開中である，横浜国立大学の森研究室で開発中の質問応答システム MinerVA-N[4]の能力を参考とする．
質問応答技術は自然言語処理技術だけでなく，情報検索技術，情報抽出技術といった多くの領域に渡る技術が使われている．質問文解析，情報検索，情報抽出，回答選択という 4 つのステップを経て質問を応答する[5].

質問文解析により，入力された質問文が何に関する質問か，何を回答すればよいかを解析する．具体的には，質問文中の『いつ』・『どんな』などの疑問詞とその周囲の単語を手がかりに，どのような固有名詞，形容詞などを回答にすればよいか(回答タイプ)を判断し，その周囲の単語をキーワードとする．

この回答タイプをもとに，新聞記事数年分，質問応答システムでいえば WEB といった知識源に対して，回答タイプと質問文にあったキーワードをもとに検索を行う．こうした知識源からの検索により，回答候補を見出し，後はそれらの重みづけにより，回答優先順位を付けて回答を行う．

こうした質問応答技術は，何らかの知識源からの検索である事から，一般的な知識を問う問題(計算問題，固有名詞を問う問題等)は現状ではほぼ回答可能と思われる．
これを踏まえると質問文の形式は，質問文に，検索をかけても意味のないキーワードのみが存在し，一般的な知識を問う質問文ではない事が条件となる．また，送信者が明快に回答できる事が必要である．

こうした条件に当てはまると考えられる，質問文の構成および回答方法として，質問文

Vol.2009-SE-164 No.14
Vol.2009-EMB-13 No.14
Vol.2009-CSEC-45 No.14
2009/5/29

の内容に中に回答が分解して存在しており，それらを組み立てることで回答するような方法である．具体例を以下に示す．

質問文例①
『い』と『ち』を続けた数字は何ですか？
回答 1(いち)

質問文例②
『かご』から『か』を消した数字は何ですか？
回答 5(ご)

数字を意味する言葉を単語として区切るなどして質問文中に分散させ，それらを集めたり(質問文①)，余計なものを削除する(質問文②)事で回答を作成する，といった質問応答形式である．

3.2 質問文構成手順について

上記に挙げた質問文は①が単語組み合わせによる回答抽出方式であり，②は単語の消去による回答抽出方式を用いている．これらの構成手順について以下に記述していく．

・単語組み合わせによる方式
ex1.『い』と『ちを続けた数字は何ですか？
(1) 回答である，『1(いち)』を１字の単語に区切る．
→ 『い』，『ち』
(2) 区切った単語の間に，関係を示す助詞『と』『に』などを加える．
→ 『い』と『ち』
(3) 区切った単語同士で，『いち』と並べられるような意味を示す動詞(繋げる，続ける，etc)，対応する助詞(『を』)を文中に入れて質問文を完成させる．
→ 『い』と『ち』を続けた数字を何ですか？

・単語の消去による方式
ex2.『かご』から，『か』を消した数字は何ですか？
(1) 回答である，『5(ご)』に，適当な単語を追加する．
→ 『か』を追加
(2) 単語の間に，助詞『と』などを加える．
→ 『か』と『ご』
(3) 区切った単語同士で，『か』を読まずに『ご』だけを示す動作を示す語(消す，抜かす，など)，動詞に対応する助詞(『から』)を文中に入れて質問文を完成させる．
→ 『かご』から『か』を消した数字
は何ですか？

例においては，1 桁の数字を回答としているが，1 桁の数字のみでは，無作為の状態でも高い確率で突破出来てしまうため，2 桁の回答にする事も考えられる．

3.3 SPIT 判別の流れ

上述してきた提案手法を用いたチューリング・テストの大まかな流れを(図 3.1)以下に示す．

(1) 発信者が電話番号を入力し，電話をかけようとする．(セッション確立要求)
(2) SIP サーバはセッション確立要求を受け，発信者に対して，提案手法によるシステムに認証要求する．
(3) 提案手法によるシステムにより，発信者へチューリング・テストを行い，その結果を SIP サーバに通知する．
(4) SIP サーバは結果によって然るべき対処を行う．（正常ならば，セッションを確立し，受信者へ着信する．SPIT ならば，切断 or 受信者に許可を求める etc）

Vol.2009-SE-164 No.14
Vol.2009-EMB-13 No.14
Vol.2009-CSEC-45 No.14
2009/5/29

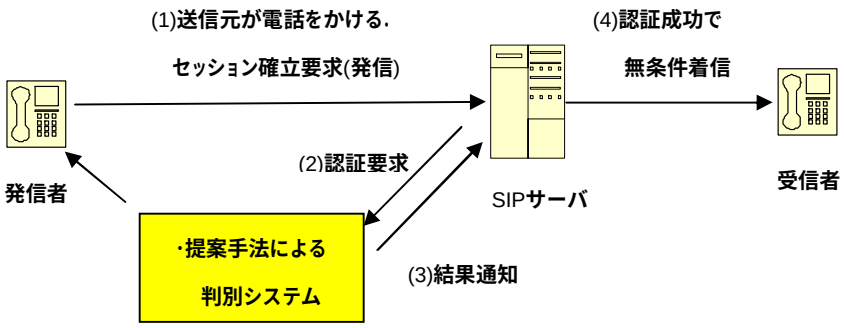

図 3.1　SPIT 判別の大まかな流れ

4.　手法評価

4.1 評価方法について

構成した質問文を質問応答システム MinerVA-N に与え，応答特性を評価する．以下のように回答の正否を評価した．

MinerVA-N は回答候補を 20 まで挙げるため，挙げられた回答候補の中に 1 つでも正解があれば正しい回答と評価する．

·質問応答システムの回答候補中に漢数字や平仮名による正解のものがあれば正しい回答と評価する．

·評価対象となる質問文は 2 つであるが，動詞や助詞などが異なる文は別個のものとして 1 つずつ評価を行う．

以上を踏まえ，質問文を 2 つ，単語の組み合わせによる方式と単語の消去による方式の 1 つずつ，全部で 16 の質問文をそれぞれ評価した．以下に質問文と回答の正解を示した．

以上を踏まえ，質問文を 2 つ，単語の組み合わせによる方式と単語の消去による方式の 1 つずつ，全部で 16 の質問文をそれぞれ評価した．以下に質問文と回答の正解を示した．

質問1.『う』と『ご』から『ご』を消して答えて下さい。　　…回答正解 5(ご)
1.1『う』と『ご』から『ご』を消して数字で答えて下さい。
1.2『う』と『ご』から『ご』を抜いて数字で答えて下さい。
1.3『う』と『ご』から『ご』を消して答えて下さい。
1.4『う』と『ご』から『ご』を抜いて答えて下さい。

質問2.『よ』と『ん』を続けて答えて下さい。…回答正解 4(よん)
2.1 『よ』と『ん』を続けて答えて下さい。
2.2 『よ』と『ん』を繋げて答えて下さい。
2.3 『よ』と『ん』を結んで答えて下さい。
2.4 『よ』と『ん』を続けて数字で答えて下さい。
2.5 『よ』と『ん』を繋げて数字で答えて下さい。
2.6 『よ』と『ん』を結んで数字で答えて下さい。
2.7 『よ』に『ん』を続けて答えて下さい。
2.8 『よ』に『ん』を繋げて答えて下さい。
2.9 『よ』に『ん』を結んで答えて下さい。
2.10『よ』に『ん』を続けて数字で答えて下さい。
2.11『よ』に『ん』を繋げて数字で答えて下さい。
2.12『よ』に『ん』を結んで数字で答えて下さい。

4.2 評価結果

各質問文を，質問応答システムを通した評価結果は以下の表 4.3～表 4.5 のようになった．

表 4.2 質問 1.1～1.4 評価結果

	質問1.1	質問12	質問1.3	質問1.4
正答数	0	0	0	0
回答数	20	20	20	20

Vol.2009-SE-164 No.14
Vol.2009-EMB-13 No.14
Vol.2009-CSEC-45 No.14
2009/5/29

表 4.3 質問 2.1～2.6 評価結果

	質問2.1	質問2.2	質問2.3	質問2.4	質問2.5	質問2.6
正答数	0	0	0	0	0	0
回答数	20	20	20	20	20	20

表 4.4 質問 2.7～2.12 評価結果

	質問2.7	質問2.8	質問2.9	質問 2.10	質問 2.11	質問 2.12
正答数	3	4	4	5	5	5
回答数	20	20	20	20	20	20

質問 1.1～1.4 で正しく回答できていたものはなく, 質問 2.1～2.9 でも回答出来ていなかった．しかし，質問 2.10～2.12 においては，正しい回答を挙げる事が出来ていた．

4.3 評価結果に対する考察

基本的に正しく回答できた質問は少なかったが，この理由は名詞の認識が質問応答システムではうまくいかなかった事が原因と考えられる．例えば，質問 1 では，『う』と『ご』を『うとご』という名詞として認識し，キーワードとして文書検索をかけたためと思われる．そのため，回答に関連がなく，意味の全く違う候補が上がったため，正しく回答する事ができなかったものと思われる．これは，質問文が様々な形で入力される公開された場にある特性上厳密に解析する事が出来ない部分もあると思われる．

正しく回答できたものは，質問 2.7～2.12 となった．この原因としては，『よにん』という名詞の認識により，『四人』というキーワードが多く検索された結果によるものと思われる．また，付録でのすべての回答候補も含めて，質問文中に『数字』を示し得るキーワードが存在するかどうかで，数字の回答候補の数がかなり違う事もわかった．
評価のまとめとしては，質問応答システムによって提案方式のような質問応答を高い確率で正答可能な条件として名詞として数値を表すような(4 人:よにん)読みの並びを含めてしまうと，質問応答システムによる正答確率が上がる事が考えられる．
その部分に気をつけていれば，人と機械による正答率の差により，SPIT の判別は可能と思われ，SPIT の判別のためのチューリング・テストとしてある程度機能していると言える．

この点に留意すれば，人と機械による正答率には大きな差があるため，SPIT の判別は可能であり，チューリング・テストとして有望と言える．

5. おわりに

本稿では, IP 電話の SPAM である SPIT を発信から通話までの過程をすべてコンピュータが自動操作すると仮定して，その判別手法を検討した．特に，SPIT 側が質問応答システムを利用して応答することを想定し, SPIT 判別上望ましいと思われる質問構成法を基本検討した．さらに現状で利用可能な質問応答システムを用いて評価し，本提案によるチューリング・テスト方式が SPIT 判別に有効である見通しを得た．

今後の課題として，本提案の応用を検討する事が挙げられる．本稿においては，質問文の方式は，単語の組み合わせによる方式と，単語の消去による方式の 2 つを提案した．この 2 つ以外にも，人には即座に回答が可能で，かつコンピュータには回答が難しい性質をもった質問文の方式を検討する事で，より SPIT 側の対応を困難にする事が可能と思われる．

また，提案した SPIT 判別手法のシステム化も今後の課題である

参考文献

1) ユールゲン クイテク・サベーリオ ニッコリーニ,サンドラ タルタレルリ・リナン シュレゲル, “ IP 電話におけるスパム防止法 ”, NEC 技報 vol.59 No.2/2006
2) ジェフ･モリガン著,宇夫陽次朗,藤田充典訳, “ SPAM の撃退 ” 1999 年 12 月 30 日，pp40-pp50
3) “音声 CAPTCHA も破られる” http://slashdot.jp/security/article.pl?sid=08/05/08/0233243
2008 年 5 月 30 日
4) 横浜国立大学 森研究室 “ MinerVA-N” http://www.forest.eis.ynu.ac.jp/cgi-bin/QA-www
5) 中島平三,外池滋生編著,大修館書店 “ 言語学への招待 ” 1994 年 3 月 1 日，pp227-pp237